REPLAYING THE EVOLUTION OF FMET: EXPERIMENTAL EVIDENCE THAT TRANSLATION INITIATION IN BACTERIA WAS INVADED BY A SELFISH GENETIC ELEMENT

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science
in Cellular and Molecular Biology

at the
University of Canterbury

by
Alannah M. Rickerby
2016

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Acknowledgements

I wish to acknowledge my supervisor Ant for providing endless guidance, support and knowledge. Thank you for allowing me to work on this project, and for putting up with me for the past two years. It has been a privilege working with you, and I look forward to working with you in the future. I also apologise in advance for the next three years of dealing with me.

I would also like to acknowledge everyone in the molecular biology lab group, who have helped endlessly with my lab work. In particular, I thank Ryan for laying all the foundations for this project, for introducing me to the lab, and explaining many of the concepts in a way I understand. Also many thanks to Alicia, for putting up with my constant questioning in the lab and at home, and having endless patience even when my queries seem like the silliest ideas. To everyone else in the lab, thank you for the endless support and entertainment. Our lab group is full of the friendliest and most helpful people I have ever met, and I am grateful for all the support and friendship I have received from you all.

To my flatmates, friends and family: I apologise for ignoring you all these past few months, and I promise to be better at replying to messages and calls from now on. Thank you all for still putting up with me, and insisting that I take a break every so often.

Abstract

The core machinery for protein synthesis is universal to cellular life. However, idiosyncrasies exist that differentiate the process of translation across the three domains (Archaea, Bacteria, Eukarya). One such example is found in bacteria, mitochondria and chloroplasts, where a formyl group is added to methionine prior to initiation of translation. Intriguingly, this formyl group is removed from the nascent polypeptide by peptide deformylase before protein production is complete, and appears to have no clear function. Despite this, formylation is essential to bacterial translation: interrupting formylation is deleterious. Such a well-conserved, and apparently deleterious, process would be expected to play an important role in bacterial translation initiation.

Previous work in our group has indicated that formylation, and the removal of the formyl group by deformylation, likely evolved from an ancient, plasmid-transmitted, toxin antitoxin system capable of post-segregational killing. These systems work by addicting cells to their presence; the toxin is more stable than the antitoxin, so if the gene pair is interrupted, the toxin is able to exert its lethal effects. A line devoid of the *def-fmt* gene pair was generated, and evolved in the absence of the genes for 1,500 generations. Despite suffering a large decrease in fitness, wildtype growth rates were observed after 1,500 generations, showing that bacteria are capable of wildtype growth in the absence of formylation.

Further to this, when *def-fmt* is reintroduced on a plasmid, a PSK phenotype was observed, with cell death occurring upon interruption of the plasmid.

We have now investigated the initial evolution of formylation in bacteria by reintroducing the *def-fmt* gene pair into a line evolved in its absence for 1,500 generations. A further 3,000-generation evolution experiment was performed with these lines. Our results indicate that an immediate reassertion of addiction has occurred. Once we reintroduce these genes into the genome, they immediately become addictive once again. We have been unable to knock-out these genes, and they also appear to out-compete cells which are not performing formylation. Whole genome sequencing has revealed a number of parallel compensatory mutations across the evolved lines, as well as a number of reversions of mutations previously observed in lines evolved in the absence of *def-fmt.* These results ultimately indicate that formylation evolved as a selfish genetic element, invading bacteria and persisting through addiction as opposed to any functional advantage.

Chapter 1
Introduction

Translation is the process by which proteins are produced, and possesses features that are highly conserved throughout the three domains: eubacteria, eukaryotes and archaea. The ribosome, a catalytic ribonucleoprotein (Ban *et al.* 2000; Rodnina *et al.* 2007), is core to the synthesis of proteins, and translation by the ribosome can be traced back to the last universal common ancestor (LUCA) (Hsiao *et al.* 2009; Hoeppner *et al.* 2012; Petrov *et al.* 2014). Many fundamental features of translation are similar among the three domains (Preiss 2016), and it is possible to divide the process into four key steps: initiation, elongation, termination and ribosome recycling (Ehrenberg 2010).

The ribosome itself possesses two subunits, each of which is slightly different across the domains of life (Melnikov *et al.* 2012). The larger subunit is responsible for the formation of peptide bonds between amino acid residues through a peptidyl transferase function (Nilsson & Nissen 2005; Trobro & Åqvist 2005), while the small subunit performs the role of decoding, and is fundamental for the process of base pairing between anti-codons encoding for an amino acid and the corresponding codon sequence on the mRNA transcript (Schluenzen *et al.* 2000). The ribosome possesses three sites: A (aminoacyl) P (peptidyl) and E (exit), which play varying roles during translation (Figure 1) (Feng *et al.* 2013).

The A site is where the charged aminoacyl tRNA enters the ribosome, while the P site is where the initiating tRNA enters the ribosome, and where the peptidyl-tRNA is located (Agrawal *et al.* 1996). Finally, the E site is where the the uncharged tRNA exits the ribosome (Sergiev *et al.* 2005).

This thesis will focus on initiation of translation, in particular, the use of formylation, a highly conserved feature in bacterial translation. In order to understand how this future is unique, I will first discuss translation initiation in all three domains of life. I will then discuss formylation in further detail. Finally, I will explain our model that this process evolved through invasion by a selfish genetic element.

Translation Initiation

Initiation of translation begins following transcription, where various initiation factors and the two ribosomal subunits come together on a mRNA transcript that has been transported to the ribosome (Bingel-Erlenmeyer *et al.* 2008). Elongation is then able to occur. This phase involves the production of a polypeptide chain, as the ribosome moves along the transcript mRNA (Schmeing & Ramakrishnan 2009; Villa *et al.* 2009). The process of aminoacylation involves the charging of the tRNA, through the amino acid being activated and then transferred to the 3'-terminal ribose of its appropriate transfer RNA (tRNA) (Arnez & Moras 1997; Banik & Nandi 2012). tRNAs and aminoacyl-tRNA synthetases match the amino

acid residues to its corresponding codon on the mRNA transcript, with rapid guanosine-5'-triphosphate (GTP) hydrolysis occurring when a correct anticodon is matched (Valle *et al.* 2003; Hershberg & Petrov 2008). This GTP hydrolysis step is irreversible and only occurs when the correct amino acid is matched, ultimately ensuring translational fidelity. The peptidyl transferase of the large ribosomal subunit acts by joining the amino acids together to create a polypeptide chain of amino acids through deacylation of a tRNA located in the P site, and transferring the peptide to the A site (Ramakrishnan 2002; Stark *et al.* 2002). Deacylated tRNA is then moved to the E site, so that the polypeptide sequence may grow (Kirillov *et al.* 2002). Termination of translation occurs when a stop codon is recognised on the mRNA transcript in the A site of the ribosome, and involves a number of release factors acting to release the polypeptide chain, and disassembly of the ribosome complex (Hirokawa *et al.* 2002; Laurberg *et al.* 2008). Both the ribosome and the mRNA transcript are then recycled for further cycles of translation (Janosi *et al.* 1998).

Despite the highly conserved protein core and general mechanisms of translation, there are variations between the domains of life that define the process as different between bacteria, eukaryotes and archaea. These oddities are interesting to study, as they aid in gaining an understanding of the evolution of early life (Hsiao *et al.* 2009), and how translation has diverged during evolution of the three domains. One such example is translation initiation – the topic of this thesis,

where the core machinery is well-conserved, yet the initiation process is undertaken in quite a different manner in each of the domains.

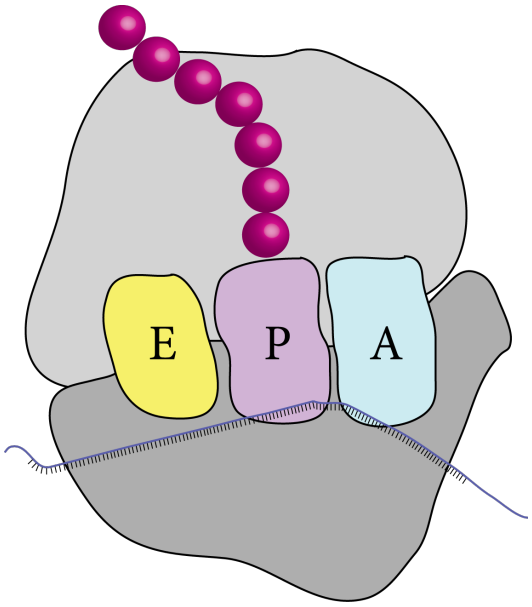


Figure 1. The ribosome is essential for the process of protein production, and is responsible for the translation of a message transcript into a polypeptide chain (pictured in purple) that ultimately becomes a protein. The ribosome possesses three sites, E (exit), P (peptidyl) and A (aminoacyl), all performing important functions during translation. The small (bottom) and large (top) subunits are also pictured in grey. From right to left, the A site is where charged aminoacyl tRNAs enter the ribosome, while the P site is both where the initiating tRNA enters the ribosome, and where the peptidyl-tRNA is located (Agrawal *et al*. 1996). The E site is where the the uncharged tRNA exits the ribosome (Sergiev *et al*. 2005).

Translation initiation in bacteria

The initiation of translation in bacteria involves initiation factors 1, 2 and 3 (IF1-3) (Figure 2) (Simonetti *et al.* 2008), and these factors all play a fundamental role in initiation (Kozak 1999). IF3 binds to the 30S subunit of the ribosome and causes dissociation of the mRNA and deacylated tRNA from the previous translation cycle, and prevents premature 70S ribosome formation (Antoun *et al.* 2006). IF1 then aids in directing the formylmethionyl-tRNA to the P-site of the ribosome by blocking the A-site (Carter *et al.* 2001). In bacteria and archaea, the mRNA contains a sequence called the Shine-Dalgarno sequence, which binds to an anti-Shine-Dalgarno sequence at the 5' end of the 16s rRNA in the 30s ribosomal subunit (Kaminishi *et al.* 2007). The interaction between the Shine-Dalgarno sequence and its complementary sequence in the ribosome aids in positioning the formylmethionine initiator tRNA at the ribosomal P-site (Antoun *et al.* 2006). When the initiation complex is complete, IF1 is bound to the A-site, IF2 sits over the A-site, the fMet-tRNA (formylmethionine-tRNA) resides in the P-site and IF3 is in the E-site, creating an appropriate conformation for translation (Simonetti *et al.* 2008; Julián *et al.* 2011). IF1 and IF3 subsequently disassociate and IF2 recruits the 50s subunit to the complex, forming the 70s ribosome (Antoun *et al.* 2003, 2006; Goyal *et al.* 2015). IF2 is then released once it positions the initiating tRNA into the correct position, and the ribosome now begins the elongation phase (Laursen *et al.* 2005).

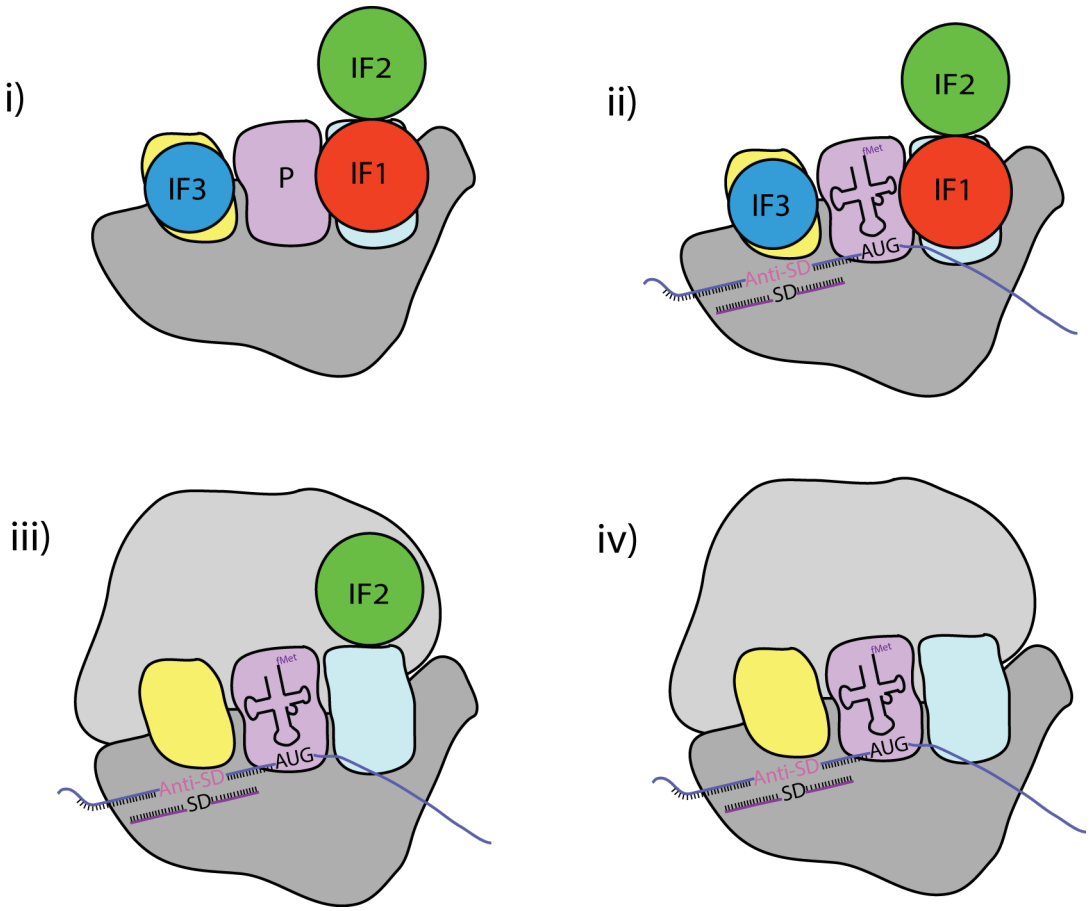


Figure 2. **Translation initiation in bacteria involves a number of initiation factors mediating the components necessary for initiation**. **i)** IF1 binds to the A site, blocking the site, while IF2 sits above the A site and IF3 sits on the E site. **ii)** IF1 directs the formylmethionyl-tRNA to the P site, while the Shine-Delgarno – anti-Shine-Delgarno interaction aids in positioning the formylmethionine initiator tRNA. **iii)** IF1 and IF3 then dissociate and IF2 recruits the 50s subunit, forming the 70s ribosome. **iv)** IF2 is released and initiation is complete. Elongation now begins.

Translation initiation in eukaryotes

While translation initiation in bacteria only involves three initiation factors, at least twelve are involved in eukaryotic translation initiation (Figure 3)(Antoun *et al.* 2003). Eukaryotic messenger RNA possesses two important post-transcriptional modifications that play a role in the initiation of translation: a polyadenylated 3' tail (poly(A) tail), made up of between 50 and 300 As, and a 7-methylguanosine cap (m^7G cap) at the 5' end (Pestova *et al.* 2007; Thoma *et al.* 2008). Both of these modifications are acted upon by various initiation factors during translation. eIF4E binds to the cap and recruits the eIF4G initiation factor (Andreou & Klostermeier 2014). eIF4G is a scaffold protein, which along with eIF4E, binds to the helicase eIF4A to form the eIF4F complex (Rogers *et al.* 2001; Oberer *et al.* 2005). The 3' end of the mRNA is acted upon by PABP, a poly(A)-binding protein, that binds to the poly(A) tail and aids in bringing together the 5' and 3' ends of the mRNA during translation (Wells *et al.* 1998).

A complex is then formed between the eukaryotic initiation factor 2 (eIF2), GTP and the initiating methionyl tRNA (Met-$tRNA_i^{Met}$), which is subsequently delivered to the 40S subunit and ribosomal P-site (Dennis 1997; Hashem *et al.* 2013). Separately to this, multiple initiation factors, eIF1, eIF1A, eIF3 and eIF5, form a preinitiation complex with the 40S ribosomal subunit, termed the 43S preinitiation complex (Gebauer & Hentze 2004; Terenin *et al.* 2016). eIF1 binds to the P-site, preventing the binding of the Met-$tRNA_i^{Met}$ in this site, and ensuring

that translation only commences when the correct sequence is identified (Lomakin *et al.* 2003; Preiss & Hentze 2003). This complex also has eIF5 bound in the P-site of the ribosome, while eIF3 is bound in the E-site and eIF1A is bound in the A-site (Passmore *et al.* 2007; Weisser *et al.* 2013). This conformation means that it is possible to bind the aforementioned mRNA complex in an open cleft on the 40S subunit (Pestova *et al.* 2007). This is mediated by an interaction between the eIF3 of the 43S preinitiation complex, and the eIF4F complex that is bound to the messenger RNA (Kolupaeva *et al.* 2005).

The 40S ribosomal subunit then moves along the mRNA in search of a Kozak sequence upstream of the 5'cap, a process referred to as scanning (Kozak 1987). Within this sequence is an AUG start codon (Hasenöhrl *et al.* 2009). When such a site is reached, the initiator methionyl tRNA recognises the appropriate sequence, displacing eIF1 from the P-site (Passmore *et al.* 2007). eIF1, eIF1A, eIF2 and eIF3 must also be displaced, as these factors prevent the binding of the 60S subunit to form the 80S ribosome (Hinnebusch 2006). This step is achieved through the involvement of eIF5, a GTP-ase activating protein and eIF5B, a GTP-ase (Passmore *et al.* 2007). Association of the 60S subunit is promoted by eIF5B, resulting in the hydrolysis of GTP bound to eIF2 by eIF5B (Lee *et al.* 2002). This causes dissociation of not only eIF2, but also the other initiation factors as well, allowing translation to occur (Kozak 1999).

Translation in eukaryotes can also be mediated without the 5' cap on the mRNA (Filbin & Kieft 2009). In these situations, the translational machinery is recruited by internal sequences within the mRNA, and may require fewer initiation factors. These sites within the RNA are termed internal ribosome entry sites (IRESs), and thousands of examples have been observed in both viral and eukaryotic RNAs (Weingarten-Gabbay *et al.* 2016).

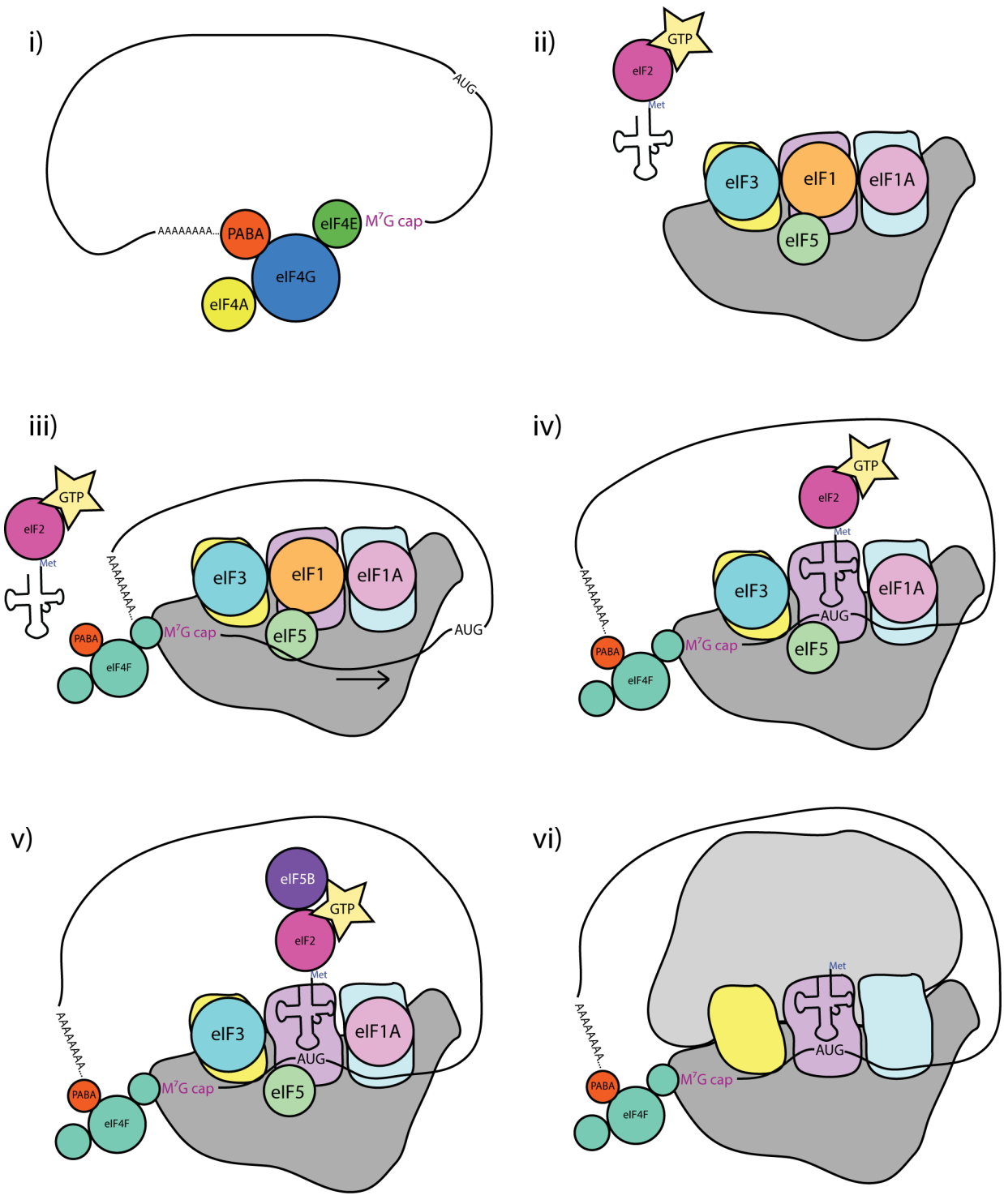

i)
AUG
AAAAAAAA...
PABA
eIF4E
M7G cap
eIF4G
eIF4A
ii)
GTP
eIF2
Met
eIF3
eIF1
eIF1A
eIF5
iii)
GTP
eIF2
Met
AAAAAAAA...
eIF3
eIF1
eIF1A
PABA
M7G cap
eIF5
AUG
eIF4F
iv)
GTP
eIF2
Met
AAAAAAAA...
eIF3
eIF1A
AUG
PABA
M7G cap
eIF5
eIF4F
v)
eIF5B
GTP
eIF2
Met
AAAAAAAA...
eIF3
eIF1A
AUG
PABA
M7G cap
eIF5
eIF4F
vi)
Met
AAAAAAAA...
AUG
PABA
M7G cap
eIF4F

Figure 3. **Translation initiation in eukaryotes is more complex than bacterial translation initiation, and involves a larger number of initiation factors**. **i)** An eukaryotic initiation factor complex, termed eIF4F binds to the m^7G cap. This is composed of the eIF4E, eIF4G and eIF4A initiation factors. PABP, a poly(A)-binding protein, also binds to the 3' poly(A) tail, and brings together the 5' and 3' ends of the mRNA. **ii)** A complex is formed between eIF2, GTP and the initiating methionyl-tRNA. Alongside this, the 40s ribosomal subunit interacts with a number of initiation factors to create the 43S preinitiation complex. eIF1 binds to the P-site, blocking binding of the initiating tRNA. eIF5 is also bound in the P site, eIF3 is bound in the E site and eIF1A is bound in the A site. **iii)** Through an interaction between eIF3 and eIF4F, the mRNA is bound in the 40s subunit. The 40S subunit then moves along the mRNA transcript, a process termed scanning, in search of an AUG-containing Kozak sequence. **iv)** When the AUG sequence is recognised by the initiating tRNA, eIF1 is displaced from the P site. **v)** eIF5, a GTP-ase activator and eIF5B, a GTP-ase, aid in assembly of the 80S ribosome through recruitment of the 60S ribosomal subunit and hydrolysis of the GTP bound to eIF2. **vi)** This ultimately causes dissociation of eIF2, eIF3, eIF5B, eIF5 and eIF1A, leaving the assembled ribosome ready for elongation.

Translation initiation in archaea

Archaeal translation initiation processes are not as well-characterised as in eukaryotes and prokaryotes, although it is thought to proceed in a manner that combines aspects of both eukaryotic and prokaryotic translation (Figure 4) (Benelli *et al.* 2003). Around ten initiation factors have been identified in archaea, significantly more than the three required for bacterial translation (Benelli *et al.* 2003). Initiation factors aIF1 and aIF1A promote the binding of a further initiation factor, aIF2, complexed with GTP, to bind in the P-site of the 30S ribosomal subunit (Hasenöhrl *et al.* 2006). aIF2 is believed to play a similar role to eIF2 during translation, forming a complex with GTP, and the initiating Met-$tRNA_i^{Met}$ (Dmitriev *et al.* 2011). This proceeds in a slightly different way to eukaryotic translation, however, with the Met-$tRNA_i^{Met}$ binding to the already ribosome-bound GTP and aIF2 instead of the ternary complex forming separate to the ribosomal subunit, as it does in eukaryotes (Hasenöhrl *et al.* 2009).

Many of the initiation factors utilised for binding to the 5' m^7G cap have not been identified in archaea, indicating that a different mechanism is perhaps utilised for this step in translation (Hasenöhrl *et al.* 2009). It is believed that this step in translation initiation may be more similar to that in bacteria, with an interaction between a Shine-Dalgarno (SD) sequence on the 5' end of the mRNA binding to a corresponding anti-SD codon in the ribosome (Sartorius-Neef & Pfeifer 2004). The binding of the aIF2•GTP:Met-$tRNA_i^{Met}$ complex in the 30S subunit results

in binding of the mRNA through the Shine-Dalgarno interaction (Benelli *et al.* 2003). Once the AUG codon of the mRNA is bound to its corresponding anti-codon in the initiating tRNA, the GTP of the aIF2•GTP:Met-tRNA$_i^{Met}$ complex is hydrolysed, resulting in the release of aIF2 and aIF1A.

The subsequent steps of translation initiation in archaea are not well understood, with the mechanism of 50S subunit binding to form the 70S ribosome not known. While the phylum is known to possess a homologue of the eukaryotic initiation factor eIF5B, its function in archaeal translation initiation is currently unknown (Hasenöhrl *et al.* 2006). In eukaryotes, eIF5B is involved with assembly of the 80S ribosome, preparing a complex capable of performing translation (Kuhle & Ficner 2014).

Some archaea are also interesting in that they undertake a process termed leaderless translation (Pfeifer *et al.* 2005). These leaderless transcripts do not possess the post-transcriptional modifications of their eukaryotic counterparts, nor do they utilise Shine-Dalgarno sequences (Benelli *et al.* 2003). Instead, leaderless transcripts are thought to involve ribosomes that are preloaded with the initiating tRNA (Londei 2005). Consistent with this, it has been observed that 70S ribosomes show a higher affinity for leaderless mRNA transcripts than 30S ribosomal subunits (O'Donnell & Janssen 2002; Moll *et al.* 2004). Leaderless translation is thought to be useful in that it may circumvent the ribosome recycling step, so that the breakdown of the large and small subunits is not

completed (Benelli *et al.* 2003). The presence of leaderless transcripts has been observed across all three domains, and would indicate that the ability to translate these transcripts is a universally conserved feature, although there is a far higher representation of this occurring within the archaea (Moll *et al.* 2002, 2004; Benelli *et al.* 2003; Li & Wang 2004).

The ribosomal machinery and structures are well-conserved across all three domains, with homologues of many of the key initiating factors being used. However, there are differences in how the process is undertaken in these domains. Examples of these differences are highlighted in eukaryotic translation, where post-transcriptional modifications result in the poly-A tail and 5'cap, both important in interacting with various initiation factors in eukaryotic translation initiation. Intriguingly also, and the topic of this thesis, is the use of a modified initiating tRNA in bacteria, where the initiating methionine is subjected to modification through the addition of a formyl group (Giglione *et al.* 2000). This is conserved across bacteria, yet is not seen in archaea or eukaryotes (Li & Chang 1995). This makes it an interesting modification to study, and might help to provide an insight into early evolution of the ribosome in bacteria.

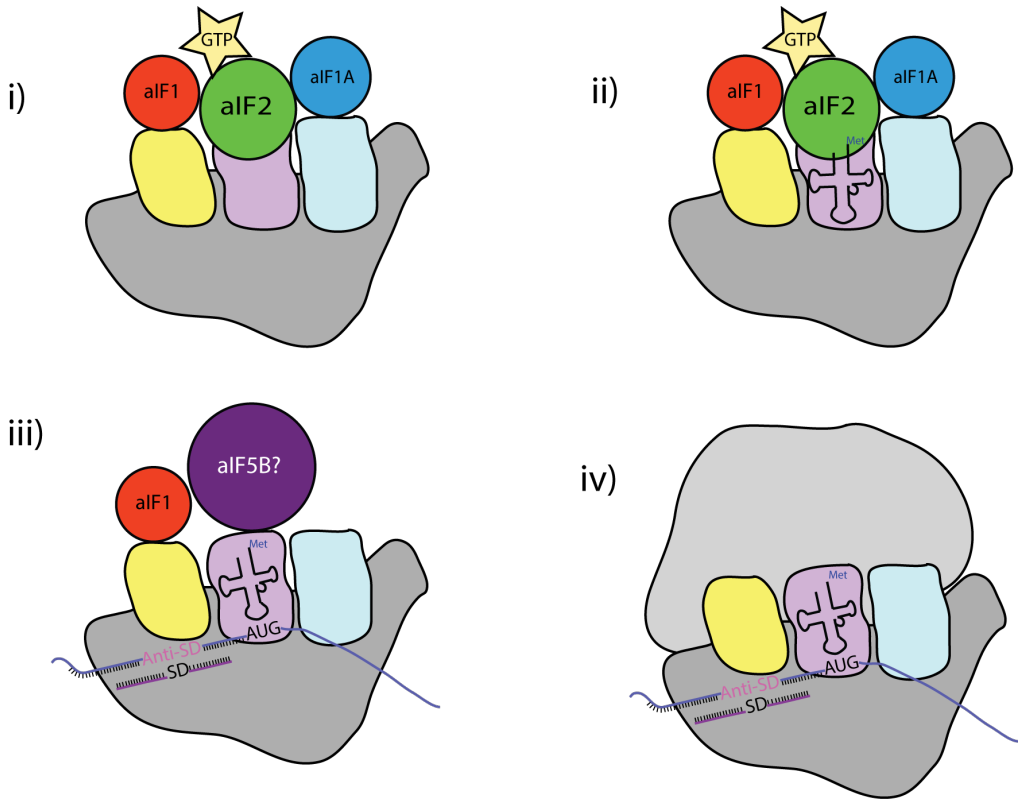


Figure 4.**Translation initiation in archaea is not as well-understood as eukaryotic or bacterial translation, but is thought to use elements of both eukaryotic and bacterial translation. i)** aIF1 and aIF1A promote binding of aIF2 complexed with GTP to the P site. **ii)** aIF2, like eIF2 is in complex with the initiator tRNA, but this complex does not form separately of the ribosome, like it does in eukaryotic translation initiation. **iii)** Binding of aIF2•GTP:Met-tRNAi^{Met} in the 30S subunit results in binding of the mRNA through Shine-Dalgarno – anti-Shine-Dalgarno interactions. Once the AUG is bound, the GTP of the aIF2 complex is hydrolysed and aIF2 and aIF1A are released. A homologue of eIF5B, termed aIF5B, may be responsible for the assembly of the 30S and 50S subunits. **iv)** The remaining initiating factors dissociate, and the fully assembled 70S ribosome is now prepared for elongation. There is

also evidence of leaderless translation occurring in archaea, where a leaderless mRNA transcript binds directly to the 70S ribosomal subunit, circumventing the need for dissociation and association of the ribosomal subunits (not shown). The exact mechanism of how this occurs is not well-understood.

Formylation of the Initiating Methionine in Bacteria

In eukaryotes and archaea, methionine is the initial residue used in the production of the polypeptide sequence, and is often removed from the nascent protein by methionine aminopeptidase (Bingel-Erlenmeyer *et al.* 2008). The removal of this methionine appears to be fundamental to cell survival, with gene replacement experiments failing to produce any viable cells (Chang *et al.* 1989; Helgren *et al.* 2016).

In bacteria, as well as the mitochondria and chloroplasts of eukaryotes (Housman *et al.* 1970; Broom *et al.* 1989), formylmethionine is used as the first residue in translational initiation (Bingel-Erlenmeyer *et al.* 2008). This is conserved across eubacteria, and involves the addition of a formyl group to the initiating methionine (Figure 5) (Mazel *et al.* 1997). The addition of the formyl group to the methionine is mediated by 10-formyltetrahydrofolate:L-methionyl-tRNAfMet N-formyltransferase (Schomburg *et al.* 2006), and is added to the charged Met-tRNAi (Schmitt *et al.* 1999). This enzyme is coded by the *fmt* gene (Mazel *et al.* 1994), and donates a formyl group from formyltetrahydrofolate to the methionine (Blanquet *et al.* 1984; Newton *et al.* 1999). The formyl group is

then subsequently removed from the nascent polypeptide by peptide deformylase as a part of the methionine cycle in bacteria (Rajagopalan *et al.* 1997; Lee *et al.* 2004; Bingel-Erlenmeyer *et al.* 2008), with the majority of mature proteins found to no longer possess *N*-formylmethionine (Clements *et al.* 2001). The deformylase gene, or *def*, forms an operon with the *fmt* gene (Mazel *et al.* 1994; Meinnel & Blanquet 1995), and deformylase appears to be a relatively unstable protein due to ferrous ion oxidation within the active site of the enzyme (Giglione *et al.* 2000; Chen *et al.* 2000). Peptide deformylase is a metalloprotease, interacting with the C-terminus of the ribosome (Bingel-Erlenmeyer *et al.* 2008). It acts by cleaving the N-formyl group so that methionine can then be removed by methionine aminopeptidase (Giglione *et al.* 2000). As a consequence of this, non-formylated methionine, alanine and serine are the most common residues observed at the N-terminus (Meinnel & Blanquet 1993).

As formylation, and subsequent removal of the formyl group by deformylase, appears to be essential to efficient bacterial growth and survival, the process has become an attractive target for antibiotic drugs (Chu *et al.* 2001; Hackbarth *et al.* 2002). Antibiotics, such as actinonin, act as inhibitors of peptide deformylase, causing cell death (Nilsson *et al.* 2006). Actinonin is a naturally occurring antimicrobial produced by an actinomycete (Gordon *et al.* 1962; Chen *et al.* 2000). When inhibition of deformylase occurs, an accumulation of toxic formylated polyptides occurs in the cell (Nilsson *et al.* 2006). This ultimately results in inhibition of bacterial growth. However, resistance has been observed

to these peptide deformylase inhibitors in both *Staphylococcus aureus* and *Escherichia coli* (Margolis *et al.* 2001). This resistance has arisen through a number of different methods. The formylation of methionine has been seen to be circumvented entirely, as well as mutations in the target gene and antibiotic efflux (Duroc *et al.* 2009). This resistance comes at a cost, however, with a decrease in growth observed in both *S. aureus* and *E. coli* (Margolis *et al.* 2001). This also highlights an ability of these bacteria to survive without formylation.

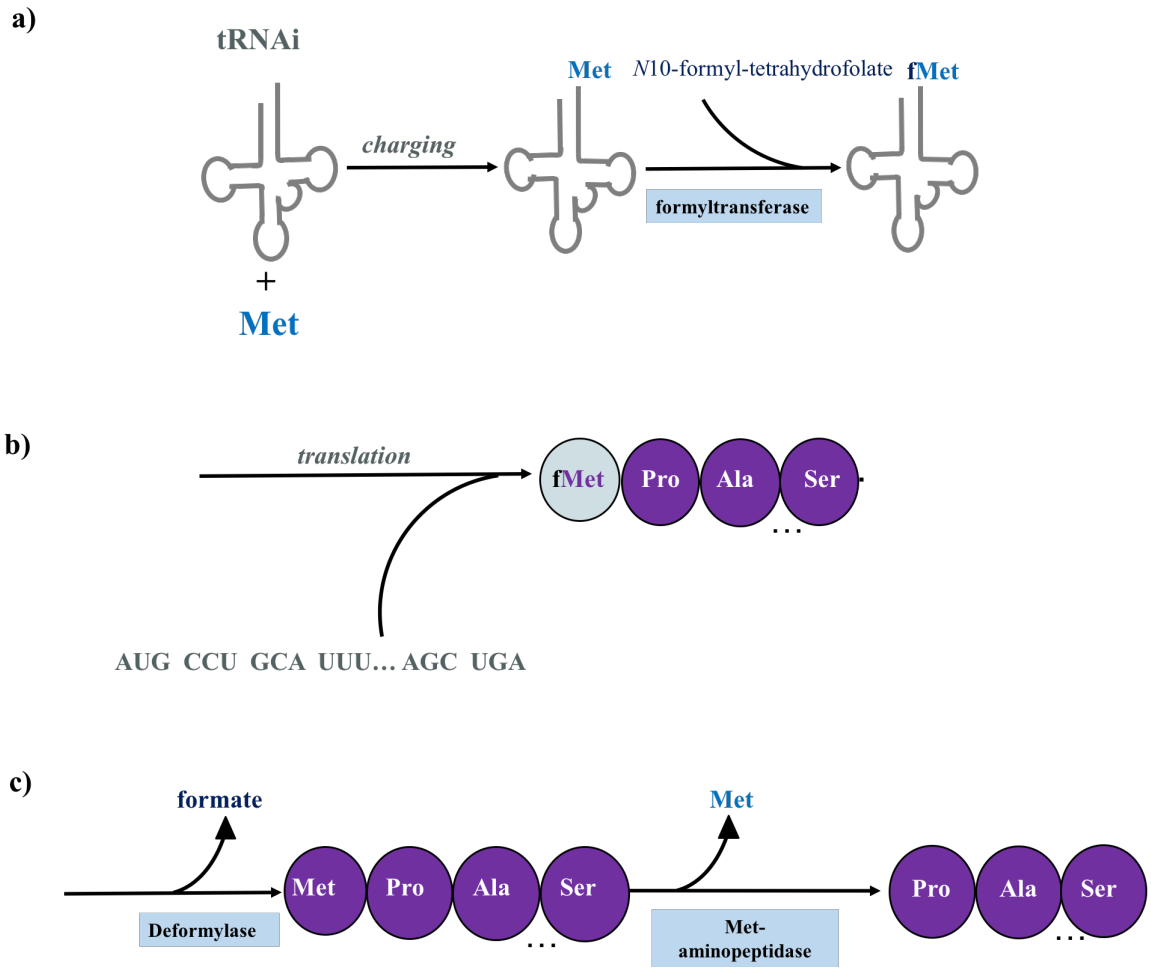


Figure 5. **Translation initiation in bacteria, and eukaryotic organelles, proceeds with a modified initiating methionine. a)** The initiating tRNAi is charged with methionine, and then is subjected to modification through the addition of a formyl group. This reaction is mediated by formyltransferase, and involves transferring a formyl group from *N*10-formyl-tetrahydrofolate to the methionine. **b)** Translation then proceeds, producing a polypeptide chain with a formylmethionine as the initiating residue. **c)** However, before a mature protein is formed, the formyl group is removed by peptide deformylase. The methionine is then also removed by methionine aminopeptidase.

Formylation is dispensable in bacteria

While the process of formylation is broadly conserved across bacteria, the use of formylation has been lost in some mollicutes, a class of parasitic bacteria that have highly reduced genomes (Grosjean *et al.* 2014). In these bacteria, translation proceeds with an unmodified Met-tRNAi^{Met}, and the genome of these mollicutes are devoid of both the *fmt* and *def* genes. Further to this, experiments have been conducted in a number of other bacterial species which show that translation without formylation is possible (Samuel *et al.* 1970; Guillon *et al.* 1992, 1996; Newton *et al.* 1999a; Li *et al.* 2000). Many of these species show varying growth effects, but knock-outs of these genes is usually associated with a severe decrease in growth rate. These experiments will be further discussed, and ultimately raises questions about the essentiality of formylation.

Guillon *et al.* (1992) established that the disruption of the Met-tRNA^{fMet} formyltransferase gene resulted in an impairment of *E. coli* cell growth. In this experiment, the formylase gene was disrupted to create mutants impaired in *fmt* function. A 1.6-fold decrease in growth rate was observed using this technique, showing the elimination of formylase to be of detriment to the bacteria, but not essential for cell viability (Guillon *et al.* 1996). Further to this, in *E. coli,* it has been noted that the loss of *def* is only possible if *fmt* has already been lost (Guillon *et al.* 1996; Margolis *et al.* 2000; Lee *et al.* 2013).

In *Pseudomonas aeruginosa* and *E. coli*, Newton *et al.* (1999a) have established that the process of formylation is expendable. The *fmt* gene was interrupted in both of these species, resulting in a severe decrease in growth in *E. coli*, but only a moderate decrease in growth rate in *P. aeruginosa*. Knocking out both of the genes (*def* and *fmt)* for the formylation process also results in impaired growth rates, although the bacteria do survive (Mazel *et al.* 1994; Yuan & White 2006).

Formylation of the initiating methionine in the mitochondria of yeast has also been found to be dispensable (Li *et al.* 2000). In this experiment, mutants of the *fmt* gene and *mis1*, a gene required for 10-formyl-THF synthesis, were found to still produce mitochondrial proteins. In these mutants, a longer lag time was observed when compared to the wild type strains.

Work undertaken by (Samuel *et al.* 1970) showed that the extracts of cells of the gram-positive bacteria, *Streptococcus faecalis* R, that had been grown without formate supplemented in their media were incapable of producing fMet-tRNAi^{fMet}. However, those extracts of cells that had been grown in the presence of formate were capable of forming the formylated methionine. The authors concluded that *S. faecalis* R was unable to formylate the Met-tRNAi^{fMet}, or that the formylmethionine was being produced by an unknown process. *S. faecalis* are unable to synthesise folic acid, a requirement of formylation (Samuel & Rabinowitz 1974). However, these bacteria are capable of growing on media that

is not supplemented with folate. Thus it was unclear how formylation could be undertaken in cells that do not possess the fundamental components to do so.

While elimination of the *fmt* gene alone can result in reduced bacterial growth, the removal of deformylase alone always results in cell death (Margolis *et al.* 2000; Lee *et al.* 2013), showing that while it is possible for bacteria to survive without the process of formylation, when peptides are formylated it is not possible for cells to survive without the removal of the formyl group by deformylase. This characteristic becomes important for our model of the evolution of formylation, where formylase is acting as a toxin to the cell and peptide deformylase as an antitoxin. This will be further discussed in this chapter.

Formylation can also be prevented in a number of organisms through the use of the antibiotic trimethoprim. Trimethoprim inhibits dihydrofolate reductase, so that formyl-tetrahydrofolate is not produced (Mazel *et al.* 1997). This ultimately prevents formylation from occurring. While treatment with trimethoprim would normally be lethal, supplementation with products produced by tetrahydrofolate, including thymidine, result in bacterial growth, despite formylation still not being possible (Mazel *et al.* 1994). Growth of cells treated with trimethoprim is reduced, but translation proceeds with an unmodified Met-tRNAi^{fMet}, demonstrating that these cells are capable of growing without the formyl group modification to the initiating methionine (Harvey 1973).

Overall, these experiments have established that formylation is not fundamental for bacterial viability, with experiments altering the *def* and *fmt* genes producing viable cells with a decreased growth rate (Margolis *et al.* 2000). This is intriguing when one considers how broadly conserved formylation is across bacteria. Formylation appears to be essential for bacterial translation, and severe growth defects or cell death are observed upon its loss, yet it is not utilised in the other domains of life. Translation initiation can also proceed in some model bacteria without formylation upon treatment with an antibiotic that prevents formylation, and some mollicutes have even lost the gene pair responsible for the process. This makes formylation interesting to study, as it is so universally conserved yet appears to be dispensable. The function of formylation is unclear, making it difficult to understand how its use has persisted. The proposed functions of formylation will now be discussed.

Function of formylation is not clear

The function of formylation in bacteria is still unclear, despite the apparent impact on cell viability and growth in the absence of the pathway (Giglione & Meinnel 2001). A number of different roles have been proposed for formylation, most of which focus on a benefit to the efficiency of the translation machinery. Petersen et al. (1976; 1978) concluded that the function of formylation was a method of protecting premature translational termination through preventing dissociation of the ribosome when translation of polycistronic mRNA is

occurring. This was argued through the observation that while the 30S ribosomal subunit showed the same affinity for Met-tRNAi^{fMet} regardless of whether it was formylated, than the 70S ribosomal subunit which showed a high affinity for fMet-tRNAi^{fMet} (Petersen *et al.* 1976a; b). Further to this, it was observed that dissociation occurred more readily in ribosomes that were using unmodified methionine when compared to formylated methionine, thus preventing premature dissociation of the ribosome (Petersen *et al.* 1978). As polycistronic messages are not seen in eukaryotes, this could explain the absence of formylation within this domain (Kozak 1999). However, the use of polycistronic messages is readily seen in archaea, which are not known to utilise formylation (Wurtzel *et al.* 2010). Later experiments have also shown that IF2 binds to the 30S ribosomal subunit, with fMet-tRNAi^{fMet} already bound to IF2 (Milón & Rodnina 2012). IF2 is known to have a higher affinity for formylated Met-tRNAi^{fMet} than its unformylated counterpart (Steiner-Mosonyi *et al.* 2004), meaning that Met-tRNAi^{fMet} might not be as likely to bind *in vivo*, compared to the results obtained from extracted ribosomes in the above experiments (Petersen *et al.* 1976a; b; 1978).

It has also been suggested that the presence of formylated methionine promotes effective translation through the prevention of side reactions by protecting the N-terminal methionine (Bingel-Erlenmeyer *et al.* 2008), but this does not explain the observation that this modification is restricted to bacteria and bacteria-derived eukaryotic organelles. If it was necessary to protect the N-terminal methionine, we would expect to see similar occurrences within the other domains of life.

Further to this, peptide deformylase removes the formyl group as the nascent chain is leaving the ribosome (Bingel-Erlenmeyer *et al.* 2008), and the methionine is removed by methionine aminopeptidase before a mature protein is formed (Chang *et al.* 1989). This would mean that any benefit of protecting the initiating methionine would be short-lived.

A third hypothesis proposed for formylation is that the presence of this formyl group allows for differentiation between elongator and initiator tRNAs, through blocking binding of formylated methionine to the elongation factor (Ramesh *et al.* 2002). This is supported by the higher affinity that IF2 possesses for fMet-tRNAi^{fMet} over Met-tRNAi^{fMet} (Steiner-Mosonyi *et al.* 2004), which would provide some sort of discrimination between initiator and elongator methionine. However, it has been observed in *P. aeruginosa*, that IF2 is capable of binding to Met-tRNAi^{fMet} nearly as efficiently as fMet-tRNAi^{fMet}. As previously mentioned, translation is able to proceed upon treatment with trimethoprim through the use of an unmodified Met-tRNAi^{fMet}. These results would indicate that IF2 is not completely discriminatory, and that translation initiation is able to proceed with both modified and unmodified Met-tRNAi^{fMet}. Finally, initiation of translation in the other two domains of life is capable of proceeding without this modification, which would indicate that formylation is not fundamental for differentiation between elongator and initiator tRNAs, if it is not necessary in archaeal or eukaryotic translation.

Function of formylation could be independent of its evolutionary origins

While a range of hypotheses exist, the function and evolutionary origins of formylation remain unclear. At this point it is important to note that the evolutionary origins of formylation, and any subsequent functional role this process may play in bacteria and organelles may not be directly connected, and it could be possible that hypothesised functions evolved at some later stage, after formylation had become established. Situations where recruitment to a new function has occurred independent of ultimate origins have been seen in a number of organisms. This includes syncytin in placental development, as well as crystallins in eye lens formation (Mi *et al.* 2000; Bateman *et al.* 2003). Synctin is a protein important in human placental morphogenesis, but has been found to have originated as a endogenous retroviral envelope protein that has been recruited to a new functional role (Mi *et al.* 2000). Decreased expression of synctins is even attributed to pre-eclampsia, showing the fundamental role these proteins play in their sequestered function (Lokossou *et al.* 2014). Such repurposing has also been observed in cystallin, a dehydrogenase that now acts as a structural protein in the eye lens of the elephant shrew (Bateman *et al.* 2003). Both these examples provide evidence of functions evolving independent of their evolutionary origins, and show that function and origins are not necessarily synonymous.

We believe that such a phenomenon may have also occurred in the case of formylation, where any subsequent functional role that formylation may serve could have evolved independent of its evolutionary origins. Our previous work has investigated the evolution of formylation in bacterial translation initiation (Catchpole 2014). We have argued that the universal formylation of methionine throughout bacteria could have evolved from a toxin-antitoxin system capable of a process called post-segregational killing (PSK). In brief, this model states that formylation could have invaded an early ancestor of bacteria on a selfish mobile genetic element such a plasmid, and persisted through its capacity for post-segregational killing. Due to this, formylation would have been able to spread through bacteria, and because of PSK, adaptation to its presence would have occurred faster than loss of the plasmid. This would have made it advantageous to no longer lose the plasmid, and at some point chromosomal integration of the genes responsible for formylation would have occurred. To better understand this model, and how formylation could have persisted in such a way, I will first review the literature of toxin-antitoxin systems.

Post-Segregational Killing Systems and Toxin-Antitoxins

Toxin-antitoxin (TA) systems are commonly gene pairs: the first of which encodes a toxin that is harmful to the cell, while the second gene encodes an antitoxin that counteracts the effects of the toxin, effectively a form of antidote (Engelberg-Kulka *et al.* 2004). If a complete TA system is not inherited by a daughter cell, the toxin is able to accumulate and exert its lethal effects (Van Melderen & De Bast 2009). This is due to the antitoxin being less stable than the toxin, resulting in the anti-toxin degrading faster than the toxin, a process termed post-segregational killing (PSK) (Anantharaman & Aravind 2003; Pandey & Gerdes 2005; Otsuka 2016). This ultimately ensures that the only cells capable of survival are those that possess the intact TA system, causing an addiction to the presence of the gene pair (Mochizuki *et al.* 2006).

TA systems are found within both bacteria and archaea, (Diago-Navarro *et al.* 2013), and while the antitoxin can be either a protein or antisense RNA, the toxin has only been observed as a protein (Holčík & Iyer 1997; Gerdes *et al.* 2005). These systems were originally found to exist on plasmids, and provide an effective method of plasmid stabilisation due to their capacity for PSK (Ogura & Hiraga 1983; Gerdes *et al.* 1986; Naito *et al.* 1995). Further experiments, however, have established that TA systems may provide a benefit when competing against other horizontally-transferred genetic elements, and that the observation of stability was due to culturing conditions (Cooper & Heinemann

2000). Horizontally-mobile plasmids that carry a PSK system are capable displacing other plasmids, and can be highly effective at spreading through a population of bacteria (Rankin *et al.* 2012; Cooper *et al.* 2010).

While it is known that plasmid-born TA systems are involved with PSK, the role of these systems encoded on bacterial chromosomes is not as clear (De Bast *et al.* 2008; Mine *et al.* 2009). A number of TA systems have been found to exist on bacterial chromosomes (Zielenkiewicz & Cegłowski 2005), and have been proposed to contribute to cellular functions, including programmed-cell death (Gerdes *et al.* 1997; Engelberg-Kulka *et al.* 2004). In *E. coli*, the *mazEF* gene pair codes for a TA system that results in programmed cell death (Engelberg-Kulka *et al.* 2004). These chromosomally-encoded TA systems have also been seen as stress-response elements that control gene expression in unfavourable conditions (Pandey & Gerdes 2005).

It is believed that PSK systems on bacterial chromosomes may have been acquired following horizontal gene transfer, with subsequent integration into the chromosome (De Bast *et al.* 2008; Van Melderen & De Bast 2009). However, initial introduction of such TA gene pairs into a cell does not necessarily correspond directly to any subsequent role that this TA gene pair may play in the cell. This idea is discussed in a more general way by Gould & Lewontin (1979), who noted that much emphasis is placed on functional importance of traits as opposed to evolutionary origins. Where traits provide some inherent benefit to

the cell, this idea is quite logical. However, in the case of traits that are either neutral or of detriment to cell fitness, it can be more difficult to argue such an adaptionist viewpoint (Lynch 2007). Instead it might be more rational to separate function and evolutionary origins (Stoltzfus 1999). TA systems often provide no fitness benefit to cells, and may instead be harmful or neutral to the cell (Cooper & Heinemann 2005; Mochizuki *et al.* 2006). However, this does not preclude a TA pair later developing a role integral to cell function, independent of its evolutionary origins, as we hypothesise for formylation.

Formylation in bacterial translation resembles known toxin-antitoxin systems

Past work by Ryan Catchpole (2014) has experimentally tested this idea that any observed function of formylation could have evolved independently of its evolutionary origins. Under this model, *def-fmt* would have arisen as a result of a PSK system that originally invaded an ancient ancestor of bacteria on a selfish genetic element. In support of this, the *def-fmt* gene pair has several characteristics possessed by known toxin-antitoxin systems, such as cell targets, stability and negative growth effects upon disruption of the gene pair. These characteristics are summarised in Table 1, and will now be discussed in further detail.

Table 1. Formylation shares a number of characteristics with known toxin-antitoxin systems.

Known post-segregational killing systems	**Formylation**
Targets include DNA replication, mRNA, protein synthesis, cell wall biosynthesis and ATP synthesis	Targets protein synthesis
Usually exist as a gene pair, one of which acts as a toxin and the other the antitoxin	Exist as the *def-fmt* gene pair in *E.coli*. the addition of the formyl group by *fmt* is toxic if not removed by *def*
Antitoxin less stable than toxin	*def,* the proposed antitoxin, is less stable than *fmt*
Disruption of gene pair results in impaired growth or cell death	Disruption of *def-fmt* results in impaired growth or cell death in a number of bacterial species
Absence of antitoxin allows toxin to accumulate and exert its effects	Absence of *def* is always fatal if *fmt* remains

Toxin-antitoxin (TA) systems are known to target DNA replication, mRNA, protein synthesis, cell wall biosynthesis and ATP synthesis (Rice & Bayles 2003; Diago-Navarro *et al.* 2013), through targeting crucial cellular processes (Yamaguchi *et al.* 2011). Formylation, like many known TA systems, targets protein synthesis through the addition of a formyl group to the initiating methionine during translation initiation (Schomburg *et al.* 2006). The phenotypic detriment to growth rate when *def-fmt* is interrupted is testament to the impact formylation is having on protein synthesis, making its target of the translational machinery (the initiating met-tRNAi) similar to those observed in other TA systems.

As aforementioned, TA systems exist as gene-pairs, where one of the genes acts as a toxin, and the second gene counteracts the toxic effects. Formylation in bacterial translation initiation is mediated by two proteins, one coded by *fmt* which adds formyl group to the initiating methionine, with peptide deformylase, coded by *def*, removing the formyl group before translation is complete. Further to this, the presence of formylated proteins is toxic to the cell if removal does not occur, and removal is generally essential for wildtype growth rates in bacteria (Mazel *et al.* 1994). The *def-fmt* gene pair resembles the appearance of many known toxin-antitoxin systems, and the toxic nature of formylation makes TA model logical to test.

Finally, as previously discussed, the antitoxin is always less stable than the toxin (Otsuka 2016). Peptide deformylase has been notoriously difficult to extract and purify due to its lability (Adams 1968; Livingston & Leder 1969; Rajagopalan *et al.* 1997). Formyltransferase, however, appears to be far more stable than peptide deformylase, with this protein being stable for months at room temperature, compared to the half life of 60 seconds for peptide deformylase (Blanquet *et al.* 1984). This labile nature allows the toxin to accumulate and exert its lethal effects if the gene-pair is interrupted (Van Melderen & De Bast 2009). The differences in stability observed between these two proteins also provides a further characteristic consistent with known toxin-antitoxin systems.

Overall, when combined, these characteristics make it logical to hypothesise formylation and subsequent deformylation as a TA system, and this hypothesis has been tested by Catchpole (2014). This has been achieved through the use of evolution experiments, creating an ideal system for testing the capacity of the *def-fmt* gene pair to act as a PSK system. These experiments will now be discussed and methods chosen will be explained in further detail. Finally, how these evolved lines were used to test PSK will be discussed.

Experimental evolution

In order to investigate the capacity of the genes responsible for formylation and deformylation of methionine to act as a toxin-antitoxin system, Catchpole (2014) generated a double knockout in *Escherichia coli*, and then serially passaged lines for ~1,500 generations, with the aim being to generate a line capable of translation without the formyl addition and removal cycle. The knockout was achieved by replacing the *def-fmt* gene pair in *E. coli* with a *nptII* cassette, conferring kanamycin resistance. Obtained knock-outs initially exhibited a severe decrease in growth rate, however had reached wildtype growth rates after 1,500 generations. Evolution experiments such as these have previously been used to monitor evolutionary changes over time, and help in answering evolutionary questions. The use of such experiments will now be further examined, in order to understand how these experiments have been previously used, and how they are relevant to our work.

Experimental evolution experiments make it possible to observe evolution in action, and have utilised a wide variety of organisms to investigate evolutionary questions (Kawecki *et al.* 2012). Many of these experiments have used bacteria due to the rapid growth rate of these organisms, as well as the large understanding about their genomes (Barrick *et al.* 2009). Evolutionary experiments have been used for investigating conditions which might be useful to a number of fields, including monitoring antibiotic resistance emergence (Gullberg *et al.* 2011),

temperature adaptation (Cullum *et al.* 2001), cellular stresses (Kawecki *et al.* 2012), and in our case, questions of evolutionary origins. These experiments make it possible to assess the effects of evolutionary changes over time through the creation of a revivable fossil record using frozen stocks (Barrick & Lenski 2013). It is also possible to control the environment in which the bacterium is grown, so changes of both phenotype and genotype can be visualised in response to changes in the environment (Barrick & Lenski 2013).

Richard Lenski has been one of the pioneers of such long-term evolution experiments, having run the evolution of twelve populations of *E. coli* for over 50,000 generations in a media where glucose is the limiting factor (Lenski *et al.* 1991; Lenski 2011). Sequencing of the genomes of these twelve *E.coli* lines at 2,000, 5,000, 10,000, 15,000, 20,000 and 40,000 generations has revealed over 600 single nucleotide polymorphisms (SNPs) and a 1.2% reduction in the genome size when compared to the ancestor (Barrick *et al.* 2009). Fitness of these lines has also increased as the bacteria have responded to growth in the glucose-limited environment, with a 3% increase in the fitness observed (Wiser *et al.* 2013). It is also possible to investigate the repeatability of evolution through these experiments as a single ancestor founded the 12 independent lines. A number of parallel changes have been observed during the duration of the long-term evolution experiment, including mutations in all 11 of their lines, and many of these have been attributed to adaptation to the culturing conditions (Wiser *et al.* 2013). These experiments have shown experimental evolution to be an effective

tool in addressing questions about adaptation in a changing cellular environment. In the case of *def-fmt,* experimental evolution has previously made it possible to investigate the ability of bacteria to respond to the absence of formylation (Catchpole 2014). Using similar conditions to those utilised by Lenski has given us a good indication of what mutations we might expect to see in our environment, as the conditions are well-studied. We are able to establish whether any mutational changes we might see can be attributed to culturing conditions (as they have been previously observed by Lenski), or whether these mutations might be specific to *def-fmt* reintroduction.

After our initial 1,500 generation-long evolution experiment, lines devoid of *def-fmt* exhibited growth rates indistinguishable from the wildtype, REL606*,* despite the lines showing an initial severe decrease in growth rate. At a genomic level, a number of parallel changes were observed between the replicate lines of *E. coli*. A number of these mutations had been observed in previous long-term culturing experiments, making it likely that these mutations were adaptations to culturing conditions, as previously discussed. Other parallel mutations observed, however, appear to have occurred only in our experimental lines. More specifically, a number of these parallel changes have been seen in genes involving the translational machinery. In all eleven lines, mutations were observed in the metZ promoter, upstream of the initiator fMet-tRNA. A GFP reporter system was developed in order to assess this further, and it was observed that this change resulted in increased expression of the initiating tRNA. It was concluded that this

mutation could be compensating for the decreased affinity of the translational machinery to unformylated methionine through increased expression.

Three mutations were also observed in *infB* in parallel in multiple lines. *infB* codes for IF2, an important initiation factor that interacts with the initiating fMet-tRNA during translation. One of these mutations has been observed in previous long-term culturing experiments, and is believed to be important to the specific culturing conditions we used. However, two of these mutations occurred in the C2 domain of IF2, the domain where fMet-tRNAi is known to bind during translation. These mutations were believed to aid in compensating for the lower affinity of IF2 for unformylated Met-tRNAi^{fMet} in bacterial translation. Overall, the mutations observed in respone to the loss of *def-fmt* show that *E. coli* are able to adapt to translation without formylation of the initiating methionine, and shows that bacteria are able to survive without this system. In order to investigate our hypothesis of *def-fmt* originally evolving as a toxin-antitoxin system, we have further studied the ability of this gene pair to act like known TA systems, which will now be reviewed.

Def-fmt exhibits a PSK phenotype when introduced on a mobile genetic element

The lines evolved devoid of *def-fmt* for 1,500 generations were subsequently used to test the capacity of *def-fmt* to act as a PSK system. This was achieved by re-introducing the *def-fmt* gene pair into the evolved lines on a temperature sensitive plasmid. Use of a temperature-sensitive plasmid made it possible to easily emulate a situation in which post-segregational killing might occur. By growing the cells at a temperature non-permissible to plasmid replication, the plasmids are forced out of the cell, replicating a situation in which a toxin-antitoxin system might be lost, resulting in PSK. Cells containing this *def-fmt* plasmid were grown alongside two controls, one possessing a known toxin-antitoxin system, *paeR7*, and one with an empty vector, so that it is possible to compare growth rates to that of a known TA system (Figure 6). Growth was monitored by measuring optical density, colony forming units (CFUs) and plasmid occupation. In the empty vector, growth and optical density continued to increase after the plasmid was forced out. In comparison, the CFU and optical density of the cells for the known PSK system was impacted severely when the plasmid was forced out. The *def-fmt* gene pair possessed growth characteristics similar to the known PSK system: that is, upon forcing the plasmid out, the growth of the cells was seriously affected. This ultimately showed that the *def-fmt* gene pair is acting consistent with what would be expected from a toxin-antitoxin system that is capable of post-segregational killing.

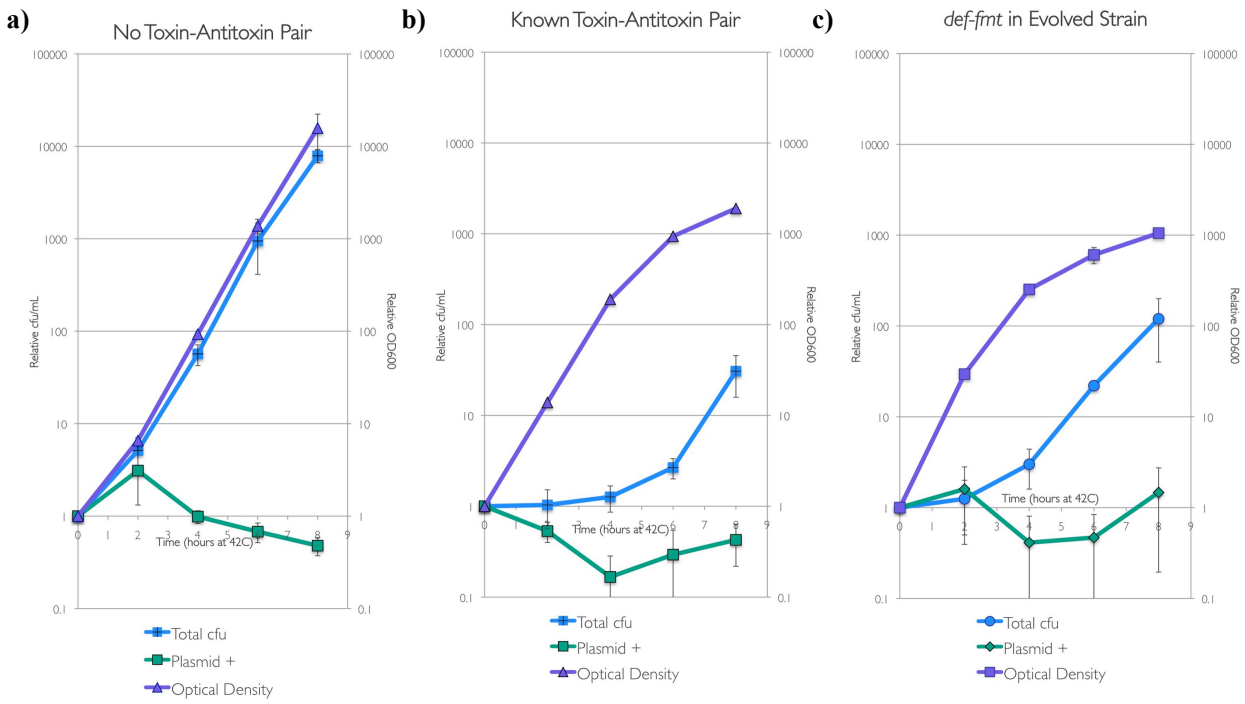


Figure 6. **The *def-fmt* gene pair exhibits a PSK phenotype**. The *def-fmt* gene pair was reintroduced into lines evolved in the absence of the gene pair for 1,500 generations on a temperature sensitive plasmid (Catchpole 2014). This plasmid makes it possible to emulate a situation where the gene pair is disrupted, allowing any PSK activity to be observed. Two controls were also used: one an empty vector, a form of negative control giving an indication of growth rates in the absence of PSK. Our second control was a known PSK system, PaeR7. Cells were grown at 37°C for 4 hours, and were then moved to 42°C, a temperature non-permissible to plasmid replication effectively disrupting the *def-fmt* gene pair. Optical density, plasmid occupancy and CFU counts were measured throughout the period of the experiment. **a)** In the absence of a PSK system (empty vector), growth is not affected. **b)** In the known PSK system, PaeR7, growth of the cells decreases when the plasmid is displaced. **c)** The *def-fmt* gene pair also shows a decrease in growth rate upon plasmid displacement, like known PSK systems. These results make the growth rate observed for our *def-fmt* lines consistent with what would be expected from a PSK system. Image from Catchpole (2014).

Summary

This thesis is intended as a continuation of previous work by Ryan Catchpole, further investigating the evolutionary origins of formylation in bacterial translation initiation. We are aiming to utilise further experimental evolution in order to investigate how formylation evolved in the bacterial genome. While we have investigated the role of *def-fmt* as a post-segregational killing system and have shown bacteria to be capable of wildtype growth rates in the absence of *def-fmt,* we have yet to investigate how this process evolved to become so conserved. In order to do this, we have taken the lines evolved devoid of the *def-fmt* gene pair, and have performed a knock-in of the *def-fmt* gene pair back onto the chromosome. We predict that we would see a number of evolutionary changes in response to *def*-fmt reintroduction after further evolution experiments, and these changes will fall under five scenarios:

1. We observe no change in genes previously mutated following deletion of *def-fmt*
2. Direct reversions of mutations that appeared following deletion of *def-fmt*
3. Mutations in genes that had not previously been mutated
4. Mutations in genes that had previously been mutated, but which are not reversions
5. No changes occur within a gene at all

Due to the dynamic nature of evolution, it is difficult to predict what exactly we might see in terms of mutational changes.

A further 100 days, or 3,000 generations, of serial culturing has now been performed in eleven replicate lines, with whole genome sequencing revealing all five possibilities occurring. We have seen a number of parallel reversions to the wildtype ancestral state, new parallel changes occurring in both new genes and genes where previous mutations have been observed, and we are seeing situations where there have been no further mutational changes in response to *def-fmt* reintroduction. We have further investigated some of these mutations, in particular a further mutation observed in *infB* (coding for initiation factor 2) in 7 of our 11 replicate lines. Three mutations have been observed previously in this gene in response to *def-fmt* loss, and this further mutation has occurred in the region where fMet-tRNAi^{fMet} is binding to the initiation factor.

Intriguingly also, is the fact that we see instant addiction to the presence of *def-fmt* once it is reintroduced, even after these lines have evolved in its absence. Upon knocking in of the *def-fmt* gene pair, we would have expected to see a 50:50 ratio of the knock-in lines to the kanamycin-cassette lines based on our method of scarless genome engineering (discussed in detail in Chapter 2). Instead, we have seen a situation were only the *def-fmt*-containing lines are observed. These results are consistent with what would be expected from a TA system, where disruption of the gene pair would result in cell growth inhibition or death. Further

to this, we have not been able to obtain *def-fmt* knock-outs when attempting to knock-out the genes from the reintroduced lines. This was attempted at both day 0 and day 80 during our culturing experiment, with both time points being unsuccessful. It would appear that addiction to *def-fmt* is being reasserted immediately upon reintroduction of the gene-pair. The results obtained in this thesis provide further evidence for formylation acting as a TA gene pair, making this hypothesis a likely explanation of how this system evolved to become so ubiquitous in bacterial translation initiation. Further to this, we are observing a situation consistent with horizontal gene transfer occurring after chromosomal integration of the *def-fmt* gene pair, providing further evidence of how this gene pair became so heavily conserved.

Chapter 2
Generation of a *def-fmt* knock-in

Introduction

Despite much of the core machinery and features of translation being universally conserved, a number of differences are seen within each domain of life. For example, the initiation factors differ between the domains, and variations in how initiation of translation is completed (see Chapter 1) (Kozak 1999). This is seen in translation initiation in bacteria, where the initiating tRNA (tRNAi) is formylated (Giglione *et al*. 2000). The initiator tRNA is firstly charged with methionine, and then the methionine is formylated (Mazel *et al*. 1997; Schmitt *et al*. 1999). This formylation reaction is mediated by 10-formyltetrahydrofolate:L-methionyl-tRNAfMet N-formyltransferase, or *fmt* (Mazel *et al*. 1994), where a formyl group is donated from formyltetrahydrofolate to the charged methionyl-tRNAi (Schomburg *et al*. 2006). Translation then proceeds, but before a mature protein is produced, the formyl group is removed from the nascent polypeptide by peptide deformylase, coded by the *def* gene (Meinnel & Blanquet 1995; Chen *et al*. 2000). A number of roles have been proposed for formylation (see Chapter 1), yet it is still unclear what role it plays. Formylation appears to be important for wildtype cell growth in bacteria, with knock-outs of the *def-fmt* gene pair showing reduced viability (Guillon *et al*. 1992). However, translation proceeds

in archaea and eukaryotes without formyl group addition (Laursen *et al*. 2005; Bingel-Erlenmeyer *et al*. 2008). This makes it difficult to explain how it may have evolved to be so universally conserved in bacteria, with a functional explanation needing to explain why it appears to be dispensable in eukaryotes and archaea.

We proposed that formylation could have become universally conserved in bacteria through mobile element spread, initially acting as a toxin-antitoxin system, resulting in a form of addiction to the gene pair, meaning that it was no longer able to be lost. Under this model, *def-fmt* would have originally invaded an ancient ancestor of bacteria on a plasmid, and post-segregational killing would have prevented its loss. As a result of this, there would have been adaptation to presence of the fMet-tRNAi, and at some point chromosomal integration of the gene pair would have occurred. This would have resulted in a situation where *def-fmt* was no longer able to be lost. Previous experiments performed by our group aimed to test this model (Catchpole 2014). In brief, we knocked the genes required for formylation (the *def-fmt* gene pair) out of *Escherichia coli* B strain REL606. These lines were then subjected to long-term culturing for 1,500 generations (DF200) in order to establish whether the bacteria were able evolve to wildtype growth rates in the absence of *def-fmt*. Despite an initial detriment to growth, after 1,500 generations, the growth rate of these bacteria was indistinguishable from the wildtype control lines. These results showed that the *def-fmt* gene pair is dispensable, and that formylation is not essential for normal

growth rates in *E. coli*. Further experiments have also shown this gene pair to act as a PSK system (see Chapter 1). While we have shown the ability of *def-fmt* to act as a PSK, we have yet to investigate the initial evolution and adaptation to the presence of formylation.

This thesis aims to investigate this initial introduction and adaptation to *def-fmt* in bacteria. To test this, we have completed further experiments into one of the evolved DF200 lines. The *def-fmt* genes were knocked back into the evolved DF200 chromosome using a scarless genome engineering method (Fehér *et al*. 2008). The production of such a knock-in allows us to replay the evolution of formylation in a line evolved to survive without the genes, answering questions of how these genes came to be so conserved. This *def-fmt* knock-in was subsequently used for further long-term culturing experiments, to assess how these lines changed in response to the reintroduction of *def-fmt*. This reintroduction will give us an indication of whether there are any fitness costs associated with the presence of *def-fmt*, and might elucidate how *def-fmt* has caused an addiction to its presence, despite no apparent benefit. For *def-fmt* to be consistent with many known toxin-antitoxin systems, we would expect to see either a negative impact on fitness, or no impact at all, and that the gene pair will not be difficult to reintroduce.

Methods

Strains and media

A single line of DF200, an *E. coli* strain obtained from Dr. Ryan Catchpole (University of Canterbury), was used for these experiments. This strain is derived from a REL606 strain (Genotype: F-, tsx-467(Am), araA230, lon-, rpsL227(strR), hsdR-, [mal+](LamS)), but has evolved in the absence of the formylase (*fmt*) and deformylase (*def*) genes for 1,500 generations, and has accumulated 58 total mutations. The REL606 strain was obtained from T. Cooper (University of Houston, Texas), and all strains were grown at 30°C or 37°C, as indicated. A description of all strains used in this thesis is provided in Figure 1.

Media used for these experiments was Davis Minimal (Difco), either supplemented with 200mg/L dextrose (DM200) or 2000 mg/L (DM2000) and 2mg/L thiamine. LB was also used, as stated (Oxoid). To make solid media, bacteriological agar was added to a concentration of 1.5% w/v (Oxoid). The following antibiotics were also used as specified at the following concentrations: streptomycin, 100µg/ml; ampicillin, 100µg/mL; chloramphenicol, 20µg/mL; kanamycin, 20µg/mL ; trimethoprim, 100µg/mL; actinonin, 100µg/mL.

Scarless knock-in protocol

A knock-in strain of the *def-fmt* gene pair was created using a scarless engineering protocol (Figure 2) (Fehér *et al*. 2008). This protocol utilises temperature-sensitive plasmids, making it possible to control the presence of the plasmids within the cells. A first donor plasmid, pST76-A, is used to introduce the gene of interest, flanked by areas 500 base pairs upstream and downstream of the area where the gene will be introduced on the chromosome. We have modified this protocol, so that the flanking regions are also flanked by restriction sites, so that it is easier to insert this region into the plasmid. The gene of interest, and the flanking regions, are cloned into a multiple cloning site on the plasmid. This plasmid is introduced into cells by transformation, and cells are grown at 30°C to enable plasmid replication to occur. The temperature-sensitive nature of these plasmids is exploited in order to integrate the gene of interest, as well as the plasmid, onto the chromosome. Streaked colonies are grown at a temperature non-permissible to plasmid replication, so that the only cells that survive on media selective for the plasmid are those that have the pST76-A plasmid integrated into the chromosome.

The cells with the integrated pST76-A plasmid are then transformed with a second helper suicide plasmid, pSTKST. This plasmid expresses I-SceI, a restriction enzyme, upon induction with chlortetracycline. This induces recombination of the region where the plasmid has integrated, by I-SceI cleaving

a recognition site contained on the pST76-A plasmid. Rec-A-mediated recombination then results in either the original wild type sequence or the replacement gene, at a ratio of 50:50.

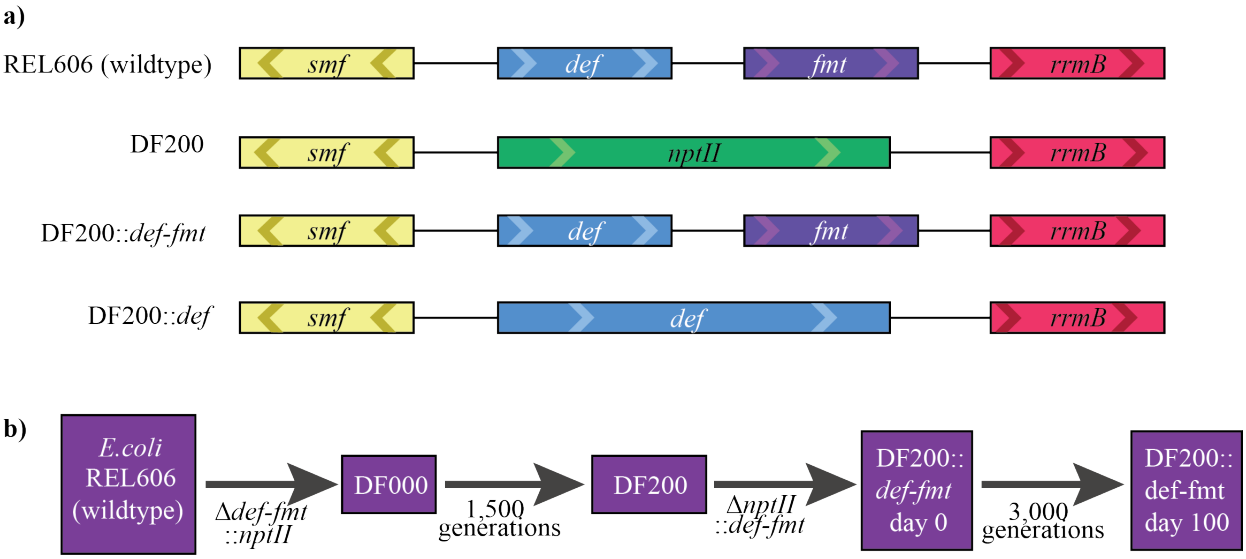


Figure 1. **Four lines of *Escherichia coli* were heavily utilised throughout this thesis. a)** Lines used throughout these experiments. REL606 *E. coli* was used for all experiments, and was the wildtype ancestor of all subsequent lines. DF200 possesses a kanamycin-resistance cassette *(nptII)* in place of the *def-fmt* gene pair, as well as 58 further mutations. DF200::*def-fmt* has all the same mutations as DF200, but has *def-fmt* in place of the kanamycin resistance cassette, while DF200::*def* possesses the *def* gene in place of the *nptII* cassette **b)** In order to obtain the aforementioned lines, a knock-out of the *def*-fmt gene pair was initially produced, giving the DF000 line. This line was then subjected to 200 days of long-term culturing (1,500 generations), giving the DF200 line. A knock-in of the *def-fmt* gene pair was next performed (DF200::*def-fmt* day 0). This line was subsequently evolved for 100 days (3,000 generations), giving DF200::*def-fmt* day 100.

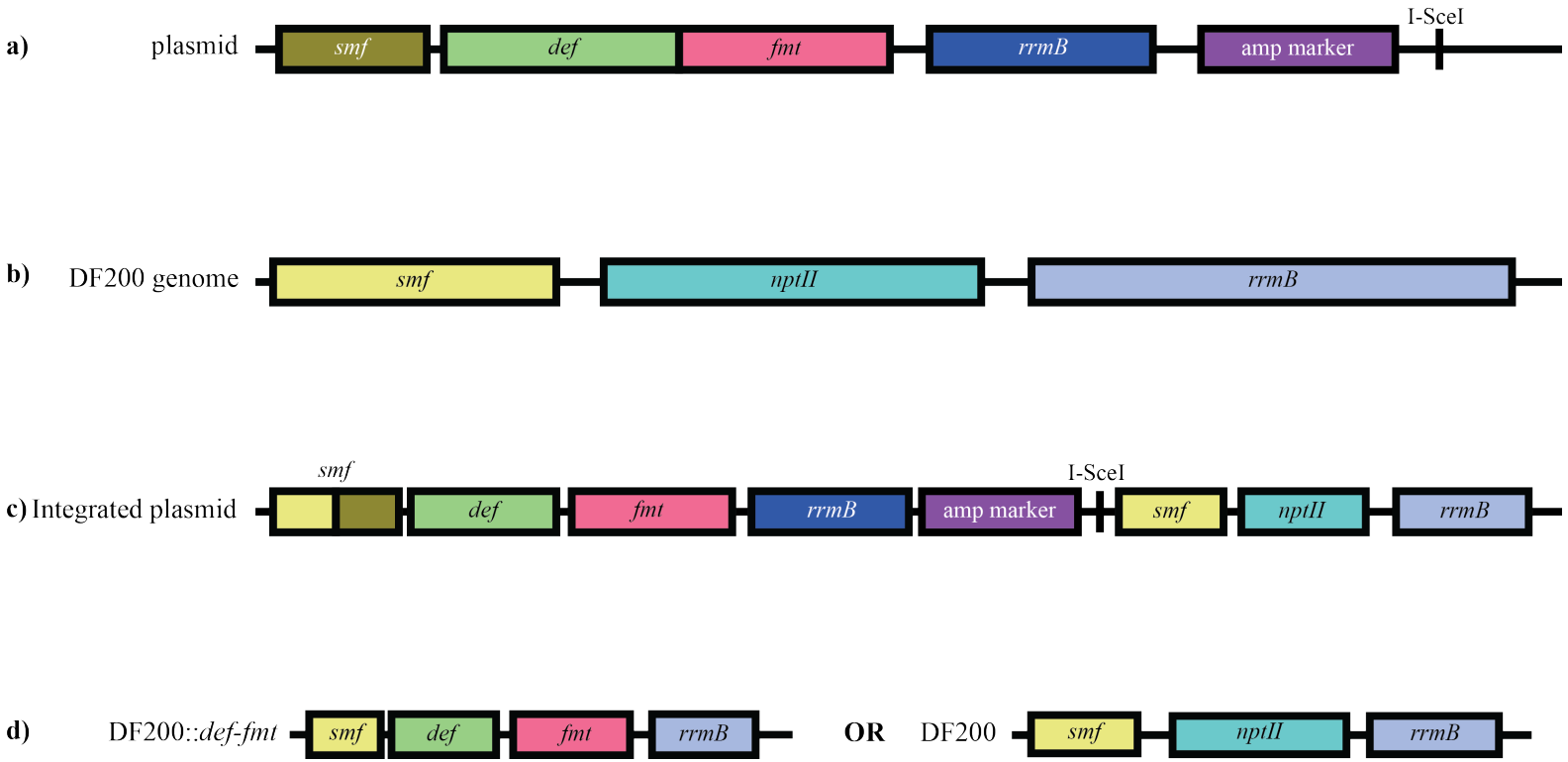


Figure 2. Scarless genome engineering was used to knock-in *def-fmt* into DF200. **a)** pST76-A::*def-fmt* harbours the *def-fmt* gene pair, as well as 500 base pairs up and downstream of the gene pair within a multiple cloning site. pST76-A also has an ampicillin marker, as well as the I-SceI cut site. This plasmid is transformed into DF200. **b)** The DF200 genome possesses a *nptII* kanamycin resistance cassette in the location where *def-fmt* resides in wildtype REL606. **c)** By growing DF200+pST76-A::*def-fmt* at a temperature non-permissible to plasmid replication, the only cells which are able to survive are those which have the pST76-A::*def-fmt* plasmid integrated into their chromosome. **d)** A second plasmid is then introduced, pSTKST, which induces recombination in the cells. pSTKST has the I-SceI restriction enzyme under the control of a chlor-tetracycline promoter. This cuts the I-SceI site on the plasmid, inducing recombination. The recombination process results in a 50:50 ratio of original DF200 harbouring a kanamycin-resistance cassette to DF200 carrying the *def-fmt* gene pair where the *nptII* cassette originally was.

Knock-in of def-fmt into DF200

The pST76-A plasmid was firstly isolated from *E. coli* MachI + pST76-A, utilising the PureLink® Quick Plasmid Miniprep Kit (Invitrogen). This plasmid was then digested with BamHI-HF (New England Biolabs) and EcoRI (Fermentas) in a sequential digest, with clean-up steps after both the first digest and second digest using the Wizard SV Gel and PCR Clean-up System (Promega), in preparation for cloning the *def-fmt* gene pair. The *def-fmt* region was amplified from the REL606 chromosome 500 base pairs upstream and downstream of the genes of interest using Phusion® High-Fidelity DNA Polymerase (New England Biolabs), with the def-500BamHI and fmt+500EcoRI primers (Table 1). These primers are flanked with the recognition sites for the restriction enzymes BamHI (highlighted in red) and EcoRI (highlighted in green), making it possible to clone the *def-fmt* region directly into a multiple cloning site in pST76-A.

The isolated *def-fmt* region was subjected to a sequential digest as aforementioned with BamHI-HF (New England Biolabs) and EcoRI (Fermentas), and was ligated with the digested pST76-A product with T4 DNA ligase (Thermo Scientific). This ligation mixture was then used to transform *E. coli* DH5α with calcium chloride and heat shock (Sambrook et al. 1989). Successful transformants were selected by plating the cells on LB supplemented with ampicillin, and were left to grow overnight at 30°C, due to the temperature-sensitive nature of pST76-A. Any colonies observed were screened using

KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems), using the T1 and T2 primers that bind over the pST76-A multiple cloning site (Table 1). pST76-A was isolated from any colonies that screened positive, and were confirmed by sequencing (Macrogen).

Table 1. Primers used for obtaining def-fmt and def knock-ins. The name of the primer used is presented on the left hand side, with its corresponding sequence next to it.

Primer Used	Primer Sequence
def-500BamHI	5′ –GGACTT GGATCC CGTAATTGTCACTCCACGCG– 3′
fmt+500EcoRI	5′ –GGACTT GAATTC TTGCCACTGCTCTGGATACG– 3′
T1	5′ –CGGAAGGATCTGAGGTTCTTATGGC– 3′
T2	5′ –CGAATTGTCGACAAGCTTGATCTGGC– 3′
fmt+678bp	5′ –CCAGACGTACAGCATCAGGG– 3′
def-808bp	5′ –GCTTCCACCACCAGTACACC– 3′
pSTKST_fwd	5′ –GGACTT GAATTC TTTCCCCGAAAAGTGCCACC – 3′
pSTKST_rev	5′ –GGACTT GAGCTC GACGAGTTCTTCTGAGCGGG – 3′
pBAD_forward	5′ –ATGCCATAGCATTTTTATCC– 3′
pBAD_reverse	5′ –GATTTAATCTGTATCAGG– 3′
def-108bp	5′ – GTACAAGCTGCTGATACTCATTAAACG– 3′
fmt+8bp	5′ – GTGGACTATCAGACCAGACGG– 3′
mutL_forward	5′ –GGCGAACCAGATTGCCGC– 3′
mutL_reverse	5′ –ATTCGGTTTTCTCGGTGCGC– 3′
hflC_forward	5′ –AAGTTCTCTGACCGTCTGCG– 3′
hflC_reverse	5′ –GTTCTCATAAGCACGCAGGC– 3′
trmB_forward	5′ –ACGGGTCAGGGAAAAAGAGC– 3′
trmB_reverse	5′ –ATTTGATGAAAACGGCCGCC– 3′
def_br	5′ –TTAAGCCCGGGCTTTCAGACG– 3′
fmt_bf	5′ –TAGTCCACTCTTCTAAGCCCGGTC– 3′
def internal fwd	5′ –CGAGACGATGTACGCAGAAGAAGG– 3′
fmt internal rev	5′ –GAACTTCTGGTTTCGCCGTGC– 3′
gstA forward	5′ – CTTTGCCGTTAACCCTAAGGG– 3′
gstA reverse	5′ – GCTGCAATGTGCTCTAACCC – 3′
def internal rev	5′ – CAGTTTGCCGACCAGGTGATCC– 3′
fmt internal fwd	5′ –CCCGGTTAAAGTTCTGGCTGAGG– 3′

The isolated pST76-A::*def-fmt* plasmid was then used to transform the evolved DF200 lines using electroporation. Transformants were selected by plating on DM2000, streptomycin, kanamycin and ampicillin agar plates and growing overnight at 30°C. Observed colonies were screened using KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems) with primers for the pST76-A::*def-fmt* plasmid: T1 and T2 (Table 1). Positive colonies were restreaked to DM2000, streptomycin, kanamycin and ampicillin and were grown at 30°C for 4 hours, then followed by 12 hours at 42°C, so that the only cells to survive would be those that had integrated the plasmid into their chromosome. These cells were then grown at 37°C for a further 16 hours. Colonies that appeared large were restreaked to DM2000, streptomycin, kanamycin and ampicillin and were grown overnight at 37°C. Any colonies observed after this were screened with the following combination of primers using KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems): T1 and fmt+678bp, and T2 and def-808bp (Table 1).

Positive colonies were then transformed with pSTKST, to proceed with the recombination step. As DF200 already possesses a kanamycin-resistance cassette in place of the *def-fmt* region, it was not possible to utilise the pSTKST plasmid with its original kanamycin resistance. The kanamycin resistance cassette in pSTKST was instead replaced with chloramphenicol resistance, to make it possible to select for the pSTKST plasmid. pSTKST-cm (pSTKSTΔ*nptII*::*cat*) was created by isolating the pSTKST plasmid from *E. coli* MachI+pSTKST using

the PureLink® Quick Plasmid Miniprep Kit (Invitrogen). pSTKST was then amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs), using the pSTKST_Fwd and pSTKST_Rev primers (Table 1).

These primers introduce recognition sites for EcoRI (highlighted in red, Table 1) and SacI (highlighted in green, Table 1). This makes it possible to easily introduce the chloramphenicol-resistance gene into the plasmid. The desired product size of ~3750bp was determined by gel electrophoresis, and the PCR product was subsequently cleaned up using the Wizard SV Gel and PCR Clean-up System (Promega), to remove products of the PCR reaction. The amplified pSTKST plasmid was next digested with SacI and EcoRI (Fermentas), in order to isolate the kanamycin-resistance marker from the plasmid. The digest was finally purified using the Wizard SV Gel and PCR Clean-up System (Promega).

The gene for chloramphenicol-resistance was then isolated from the plasmid pBSL181 in *E. coli* S17. A plasmid isolation was firstly performed using the PureLink® Quick Plasmid Miniprep Kit (Invitrogen), with the purified plasmid being subsequently digested using EcoRI and SacI (Fermentas). The digested plasmid was then subjected to gel electrophoresis, and the expected band size of ~1kb (the chloramphenicol-resistance gene) was cut from the gel. The gel fragment was purified using the Wizard SV Gel and PCR Clean-up System (Promega). This fragment was subsequently ligated with the isolated pSTKST fragment using T4 DNA ligase (Thermo Scientific), and used to transform *E. coli*

DH5α cells using calcium chloride and heat shock (Sambrook *et al*. 1989). Transformants were selected by plating on LB and chloramphenicol agar, and left to grow overnight at 30°C. Colonies observed were subjected to restriction digest mapping using SalI (Thermo Scientific) and HindIII (Fermentas) in order to confirm the correct pSTKST-cm plasmid, as pSTKST-cm was expected to possess more restriction sites than pSTKST alone. pSTKST-cm was then isolated from positive colonies using the PureLink® Quick Plasmid Miniprep Kit (Invitrogen) and confirmed by sequencing (Macrogen).

DF200+pST76-A::*def-fmt* was transformed with pSTKST-cm using electroporation (Becker & Guarente 1991). Transformants were selected for by plating on DM2000 with streptomycin, kanamycin, ampicillin and chloramphenicol, and were grown overnight at 30°C.

pSTKST-gm was created by taking the aforementioned pSTKST product digested with EcoRI and SacI (Fermentas) and ligating this with the gentamicin-resistance gene isolated from S17+pBSL182. The pBSL182 plasmid was isolated from S17 using the PureLink® Quick Plasmid Miniprep Kit (Invitrogen). The isolated plasmid was then digested with SacI and EcoRI (Fermentas), and was subjected to gel electrophoresis. The ~900bp band corresponding to the gentamicin resistance marker was cut from the gel and was purified using the Wizard SV Gel and PCR Clean-up System (Promega). This isolated gene

fragment was next ligated using T4 DNA ligase (Thermo Scientific), and the ligation mixture was used to transform *E. coli* DH5α.

The plasmid pBAD-I-SceI was obtained from Ryan Catchpole. This plasmid, while not temperature-sensitive, contains the I-SceI restriction enzyme under control of an arabinose promoter, and also possesses the same chloramphenicol-resistance marker as pSTKST-cm. This plasmid was used to transform DF200+pST76-A::*def-fmt* using electroporation. Transformants were selected by plating on solid DM2000 media, with streptomycin, kanamycin, ampicillin and chloramphenicol and were left to grow overnight at 37°C. Observed colonies were PCR screened with primers for the integrated pST76-A::*def-fmt* plasmid as well as primers for the pBAD-I-SceI plasmid. For the integrated pST76-A::*def-fmt* plasmid, the T1 and fmt+678bp, and T2 and def-808bp primers were used. For the second, pBAD-I-SceI plasmid, the pBAD_forward and pBAD_reverse primers were used (Table 1).

Colonies that yielded the correct PCR product were restreaked to DM2000 agar with streptomycin, kanamycin, ampicillin and chloramphenicol, and were grown overnight at 37°C. Single colonies were then used for the induction step. As REL606 (the original ancestor of the DF200 line) is ara-, a single point mutation means that REL606 genotypes are not able to utilise arabinose (Lenski *et al*. 1991). Furthermore, arabinose is known to inhibit glucose metabolism in *E. coli* (Koirala *et al*. 2016), making ideal conditions for the induction of recombination

uncertain. Because of this, a number of conditions were tested when inducing recombination, to ensure that the cells still grew in the presence of arabinose. The following conditions were therefore used to induce I-SceI so that recombination could occur:

1. DM with 0.2% arabinose, supplemented with 2mg/mL thiamine (no glucose)
2. DM with 0.2% arabinose and 0.02% glucose, supplemented with 2mg/mL thiamine
3. LB with 0.2% arabinose

After induction in liquid media, cultures with observed growth were then plated on corresponding solid media, as listed in the table below (Table 2). Two dilutions of each induction were plated (10^4 and 10^5), and were left to grow overnight at 37°C. Observed colonies were then screened using KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems), using primers that span the region where recombination should have occurred: def-808bp and fmt+678bp (Table 1).

Table 2. Media used for induction of recombination, listed with its corresponding solid media.

Induction media	Solid media
DM + glucose + arabinose	DM2000 agar + 0.2% arabinose, streptomycin and chloramphenicol
LB + arabinose	LB agar + 0.2% arabinose, streptomycin and chloramphenicol

Positive colonies were digested with SphI (New England Biolabs). The kanamycin-resistance cassette originally used to replace the *def-fmt* gene pair is indistinguishable from the *def-fmt* knock-in when subjected to gel electrophoresis. Because of this, digestion with SphI is necessary to distinguish between the two products. If recombination is successful, roughly half of the colonies would be expected to cut with the restriction enzyme, indicating that these colonies still carry the kanamycin-resistance cassette. The other half should remain uncut, and these colonies should possess the desired knock-in *def-fmt* sequence. Any colonies that screened positive for the *def-fmt* region were grid plated onto the following solid media:

1. DM2000 with kanamycin. Growth on this plate would indicate the kanamycin-resistance cassette is still present.
2. DM2000 with ampicillin. Growth would suggest that recombination had not been successful, and that the pST76-A plasmid was still present.
3. DM2000 with actinonin. Actinonin should kill any cells that are formylation competent, and no growth here would indicate that the knock-in was successful.
4. DM2000 with streptomycin and chloramphenicol. Growth here, when compared to growth on the other plates should indicate successful knock-ins.

Any colonies that appeared to possess the correct knock-in sequence were confirmed by PCR with Phusion® High-Fidelity DNA Polymerase (New England Biolabs), using the def-808bp and fmt+678bp, and def-108bp and fmt+8bp primers (Table 1), and were sequenced (Macrogen). Primers were also used that made it possible to distinguish DF200 from REL606. These primers are designed in genes where mutations were observed after the original long-term culturing experiment. This should confirm that I had obtained a knock-in in the correct strain of *E. coli*, and were in the *mutL, hflC* and *trmB* genes (Table 1). The amplified PCR products were then purified using Wizard SV Gel and PCR Clean-up System (Promega), and were confirmed by sequencing (Macrogen).

Any colonies that were confirmed to have the correct knock-in sequence, as well as having the correct known sequence in the *mutL, hflC* and *trmB* genes, were then grown at a 1:1,000,000 dilutions in DM2000 and streptomycin in an attempt to eliminate the pBAD-I-SceI plasmid from the culture. The diluted cultures were grown overnight at 37°C, and were subsequently plated at 10^5 and 10^6 dilutions on DM2000 solid media supplemented with streptomycin. Plates were grown at 37°C overnight, and were then grid plated to DM2000 with streptomycin and DM2000 with chloramphenicol. Any colonies that grew on streptomycin but not chloramphenicol were confirmed to have lost the pBAD-I-SceI plasmid. These colonies were used for subsequent long-term evolution experiments.

Knock-in of def into DF200

As a control for the subsequent long-term culturing experiments, a second knock-in was performed with solely *def*. This acts as control in subsequent whole genome sequencing, so that all changes seen can be attributed solely to the presence of formylation in the cell. The knock-in was performed using the scarless method, as mentioned above, however, a slightly different method was used for initially cloning pST76-A::*def*. pST76-A::*def* was cloned by amplifying the two regions flanking the *fmt* gene in order to exclude this from the knock-in, with the def-808bp and def_br, and fmt_bf and fmt+678bp primers (Table 1). A successful PCR was confirmed by gel electrophoresis. The two amplified products were then combined, and subjected to further amplification with the def-500BamHI and fmt+500EcoRI primers, so that the two previously amplified products were joined together excluding the *fmt* gene.

Following isolation of the *def* gene, the aforementioned knock-in protocol for the *def-fmt* gene pair was followed exactly.

RT-PCR

RNA was extracted from confirmed knock-ins, as well as DF200 and REL606 controls using a hot phenol protocol modified from a protocol used for *Saccharomyces cerevisiae* RNA extraction (Schmitt *et al*. 1990). Quality of

extracted RNA was confirmed by gel electrophoresis, and quantified by the Nanodrop 1000 Spectrophotometer. RNA was diluted to 200ng/µl, and a DNase treatment was completed using the TURBO DNA-*free*™ Kit (Thermo Scientific), as per the manufactures' specifications. RT-PCR was then performed using SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase (Thermo Scientific). This kit makes it possible to complete the cDNA synthesis as well as the subsequent PCR in a single PCR run, with an initial cDNA synthesis step at 55°C for 35 minutes, followed by a PCR of 40 cycles. The RT-PCR was performed using primers internal for the *def-fmt* gene pair, as well as primers for *gstA,* a housekeeping gene, as a control (Table 1). For the single *def* knock-in, different combinations of primers were used alongside the *gstA* primers in order to confirm that *def* was producing a transcript, while *fmt* was not. These included forward and reverse primers for both the *def* and *fmt* genes to confirm that a transcript was being produced by *def*, but not by *fmt* (Table 1).

Northern Blot

To confirm that the formyl group is once again being added to the initiating methionine during translation, polyacrylamide gel electrophoresis (PAGE) was utilised, followed by northern blot. RNA was isolated under acidic conditions in mid-to-late log phase, as per (Köhrer & RajBhandary 2008). Extracted total RNA was then subjected to deacylation treatments. Non-formylated aminoacylated tRNA in the total RNA was deacylated with 10mM $CuSO_4$, 0.1M sodium acetate pH5.0 for 1 hour at 37°C. Formylated total RNA was deacylated using 0.1M Tris-HCl, pH9.5 at 37°C for one hour. One further tube of RNA did not receive the deacylation treatment. Electrophoresis was then used to separate around 1000 ng of total RNA for 36 hours at 300V in a 4°C room using a 6.5% polyacrylamide gel (19:1 acrylamide:bisacrylamide), containing 8M urea and 0.1M sodium acetate (pH5.0). Two 10cm portions of the gel between the xylene cyanol and bromophenol blue bands were transferred to a Hybond-N+ membrane (GE Healthcare Life Sciences) at 10V for 16 hours using 4x TE buffer (40mM Tris-Hcl, 4mM EDTA, pH 8.0). The membranes were then subsequently baked for 2 hours at 80°C, and were soaked in prehybridisation buffer (6x SSC, 10x Denhardt's solution, 0.5% SDS) at 42°C for 6 hours. The membranes were then probed with a 5′ Digoxigenin (NHS-ester) labeled synthetic DNA probe (Integrated DNA Technologies) that corresponded to bases 20-45 of the *E. coli* initiator tRNA for 16 hours at 50°C:

5′ – DIG-NHS-CTTCGGGTTATGAGCCCGACGAGCTA – 3′

Membranes were then washed with 6x SSC twice (900mM NaCl, 90mM trisodium citrate, pH 7.0). Further washing steps in 4x SSC and 2x SSC were then carried out. The DNA probe was then detected using Anti-Digoxigenin-AP Fab fragments (Roche) and CPD-*Star* (Roche), as per manufacturers' instructions. Chemiluminescence was detected using a G:Box gel documentation system (Syngene).

Due to issues with the gel, a number of steps in this northern blot were tweaked, in an attempt to optimize the visualization of the initiator tRNA. The concentration of RNA loaded into the gel was increased, beginning at 1µg, rising to approximately 10µg through the construction of new combs that allowed loading of more sample, to ensure that enough RNA was being loaded into the gel. Several methods of RNA extraction were also utilized in order to optimize the quality of extracted RNA. Alongside the aforementioned phenol extraction protocol (Köhrer & RajBhandary 2008), RNA was also extracted using TRIzol® LS Reagent (Life Technologies), as per the manufacturers recommendations. A hot phenol extraction method was also used, following the methods of (Schmitt *et al.* 1990). Smaller pre-cast 6% TBE gels (Invitrogen) were run, and were probed as aforementioned in order to establish whether it was the gel itself proving to be the issue, or whether it was the method used in probing that was resulting in issues. Using these smaller gels, optimizing the hybridisation

temperature and length of hybridisation time was also explored to ensure that the probing step was optimal. Smaller versions of the larger gel were also cast, to investigate whether it was the length of time the gel was running that was causing issues. In order to ensure that no contamination was affecting visualization of the gel, all steps were also repeated using DEPC-treated (Sigma-Aldrich) solutions.

Investigation of knock-in ratios

Overnight cultures of DF200 + pST67-A::*def-fmt* + pBAD-I-SceI were grown at 37°C in LB supplemented with ampicillin, chloramphenicol, kanamycin, streptomycin, trimethoprim and thymidine. Control cultures of DF200 + pST76-A::*def-fmt* +pBAD-I-SceI and DF200 + pST76-A::*def* + pBAD-I-SceI were also grown in LB supplemented with ampicillin, chloramphenicol, kanamycin and streptomycin. Induction of recombination was then repeated in the presence of arabinose. For the trimethoprim treatment, cells were grown overnight in LB supplemented with 0.2% arabinose, chloramphenicol, streptomycin, trimethoprim and thymidine. The two control lines were grown in the same media, excluding the trimethoprim and thymidine. Cultures were grown overnight at 37°C, and were then plated on their corresponding solid media at a 10^5 and 10^6 dilutions. Observed colonies were then screened using the KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems), using the def-808bp and fmt+678bp primers (Table 1).

For the single *def* knock-in, it is possible to observe the difference between the knock-in and the original sequence solely with PCR. However, for the two *def-fmt* knock-in cultures being tested, it was necessary to digest the amplified PCR product in order to determine whether a knock-in had been obtained. This is due to the the *def-fmt* gene pair and the *nptII* kanamycin-resistance cassette yielding a band of the same size. Positive clones for these knock-ins were digested using SphI (New England Biolabs) for 1 hour at 37°C, and were then run on a gel to determine whether a knock-in had been obtained.

Results

Knock-in into DF200 yielded a def-fmt-containing DF200 line

To reintroduce the *def-fmt* gene pair into a line of *E. coli* that has evolved in the absence of these two genes, a scarless genome editing method was used. In order to do this, the *def-fmt* gene pair was cloned into the multiple cloning site of pST76-A. The sequence of this plasmid was confirmed by sanger sequencing (Figure 3a). This confirmed plasmid was then used to transform DF200. Successful transformants were subsequently grown at a temperature non-permissible to plasmid replication, so that the only cells to survive would be the ones with the plasmid integrated into the chromosome. A second plasmid, pSTKST-cm was then created by cloning a chloramphenicol resistance marker in place of a kanamycin resistance cassette in pSTKST. The success of this cloning was confirmed by sequencing (Figure 3b), and this plasmid was used to transform DF200+pST76-A::*def-fmt*. However, despite success in transforming this plasmid into both DH5α and REL606 strains of *E. coli,* no successful transformants were obtained in DF200. Because of this, a second pSTKST plasmid was created, this time containing a gentamicin-resistance marker. However, the gentamicin was found to be ineffective, as DF200 devoid of any plasmid conferring resistance to this antibiotic also grew in the presence of this antibiotic.

DF200+pST76-A::*def-fmt* was instead transformed with pBAD-I-SceI, a plasmid which is able to induce recombination in the same way as pSTKST, but under an arabinose promoter. This transformation was deemed to be far more successful, with many transformants obtained. Arabinose was then used to induce recombination in the successful transformants. Those transformants that possessed the correct product when screened, were then confirmed by sequencing. Sequencing across the *def-fmt* region confirmed that the correct knock-in region had been obtained (Figure 4). Further sequencing in the *mutL*, *hflC* and *trmB* regions confirmed that a knock-in had been obtained (Figure 5).

Knock-in of def produced a DF200 line harbouring the def gene

A single knock-in of the *def* gene into the evolved DF200 line was generated to act as a control for further long-term evolution experiments. This was undertaken in the aforementioned DF200 line, utilising the modified scarless genome engineering protocol (see methods). The *def* region was firstly cloned into the pST76-A plasmid within the multiple cloning site. The correct plasmid size was confirmed by screening the cells with PCR, and this plasmid was extracted then used to transform DF200. Cells that screened positive were then subjected to an integration step, where growing the cells at a temperature non-permissible to plasmid replication resulted in the only cells surviving being those which had integrated the pST76-A::*def* plasmid into their chromosome. Cells that screened

positive were then transformed with a second plasmid, pBAD-I-SceI. Observed transformants were confirmed by PCR colony screening, and recombination was then induced in these cells with arabinose. These cells were screened by PCR, and those which possessed the correct region were confirmed by sequencing (Figure 6). The knock-in was confirmed by sequencing regions which were known to be different in the evolved DF200 lines compared to the original ancestor strain, REL606. This involved sequencing the *mutL*, *hflC* and *trmB* regions of the knock-in line, and comparing these sequences to the original REL606 wildtype, as well as the evolved DF200 genome (Figure 7). From this, it was concluded that a knock-in of *def* in the DF200 background had been obtained.

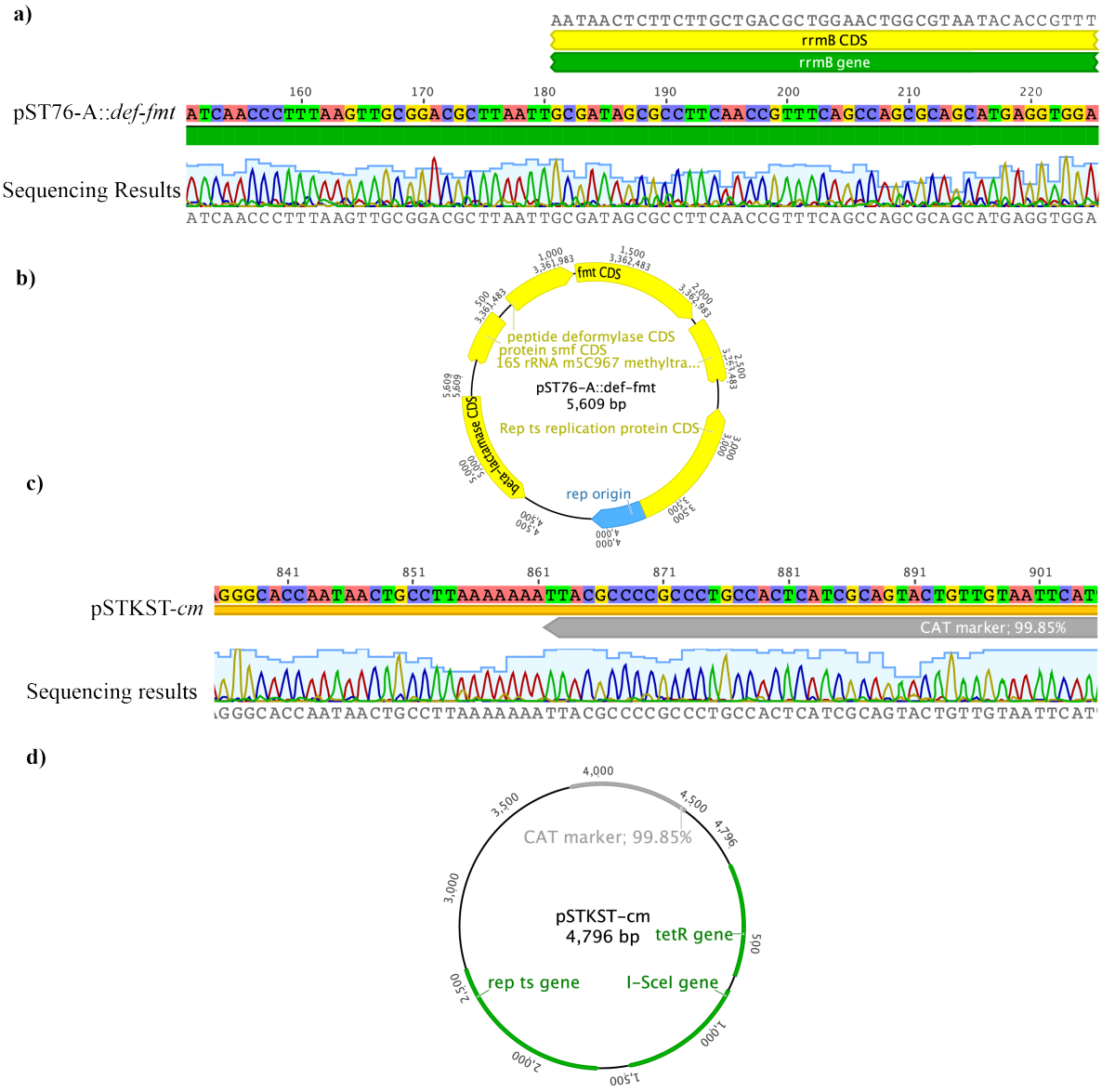


Figure 3. **Sequencing of pST76-A and pSTKST show the correct cloning of pST76-A::*def-fmt* and pSTKST-cm. a)** Sequencing of pST76-A over the multiple cloning site shows *def-fmt* correctly inserted into the pST76-A plasmid, with 500-bp upstream of the *def-fmt* gene pair, *rrmB*, shown within the plasmid. **b)** This has produced the pST76-A::*def-fmt* plasmid containing the *def-fmt* gene pair to be used for reintroduction of the *def-fmt* gene pair. **c)** Sequencing using pSTKST primers reveals introduction of chloramphenicol-resistance marker, replacing the original kanamycin-resistance cassette. **d)** This has given the pSTKST-cm plasmid, containing chloramphenicol resistance and expressing I-SceI under the chlortetracycline promotor.

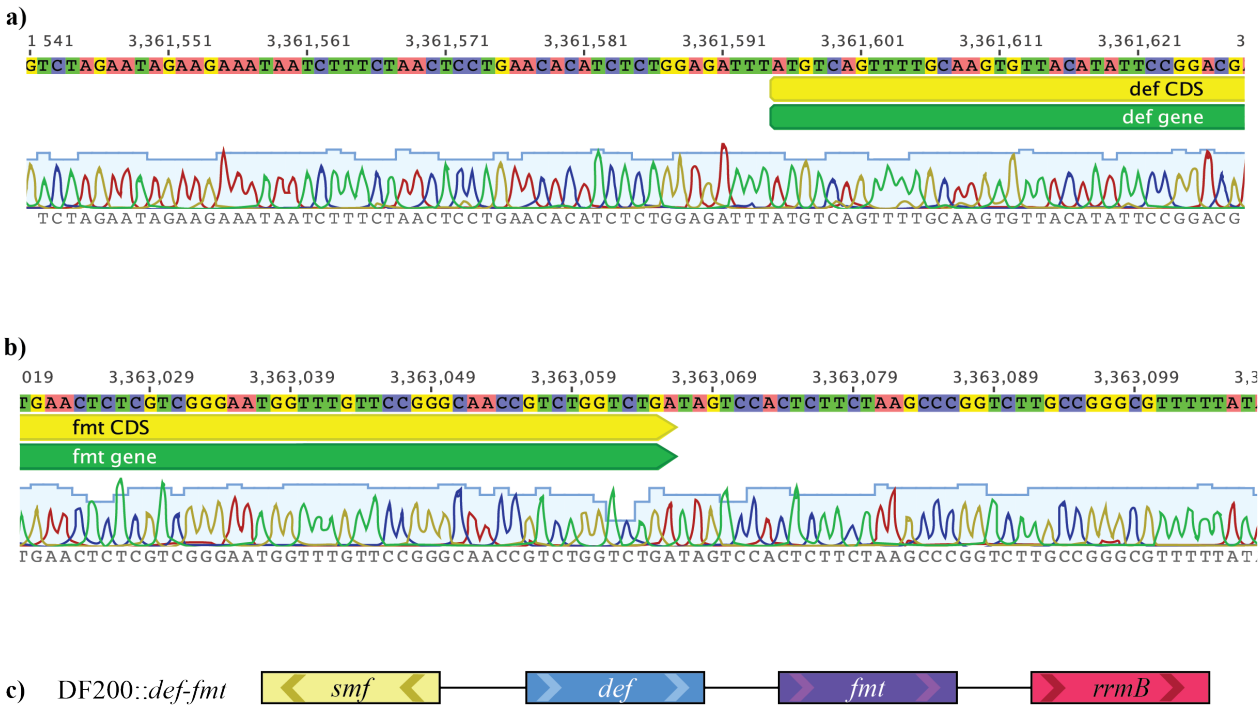


Figure 4. **Sequencing across the *def-fmt* region reveals a successful knock-in of *def-fmt* into DF200.** Primers 808 base pairs and 180 base pairs upstream, and 8 base pairs and 678 base pairs downstream were used to insure *def-fmt* has been reintroduced at the correct place in the DF200 chromosome. **a)** Sequencing across the junction where *def* should have been knocked in reveals that the gene has been introduced in the correct place. **b)** Likewise, sequencing across the junction where *fmt* was introduced shows that the correct knock-in had been obtained. **c)** Sequencing across the whole *def-fmt* region revealed that the *def-fmt* gene pair had been introduced in the correct place, producing the DF200::*def-fmt* line.

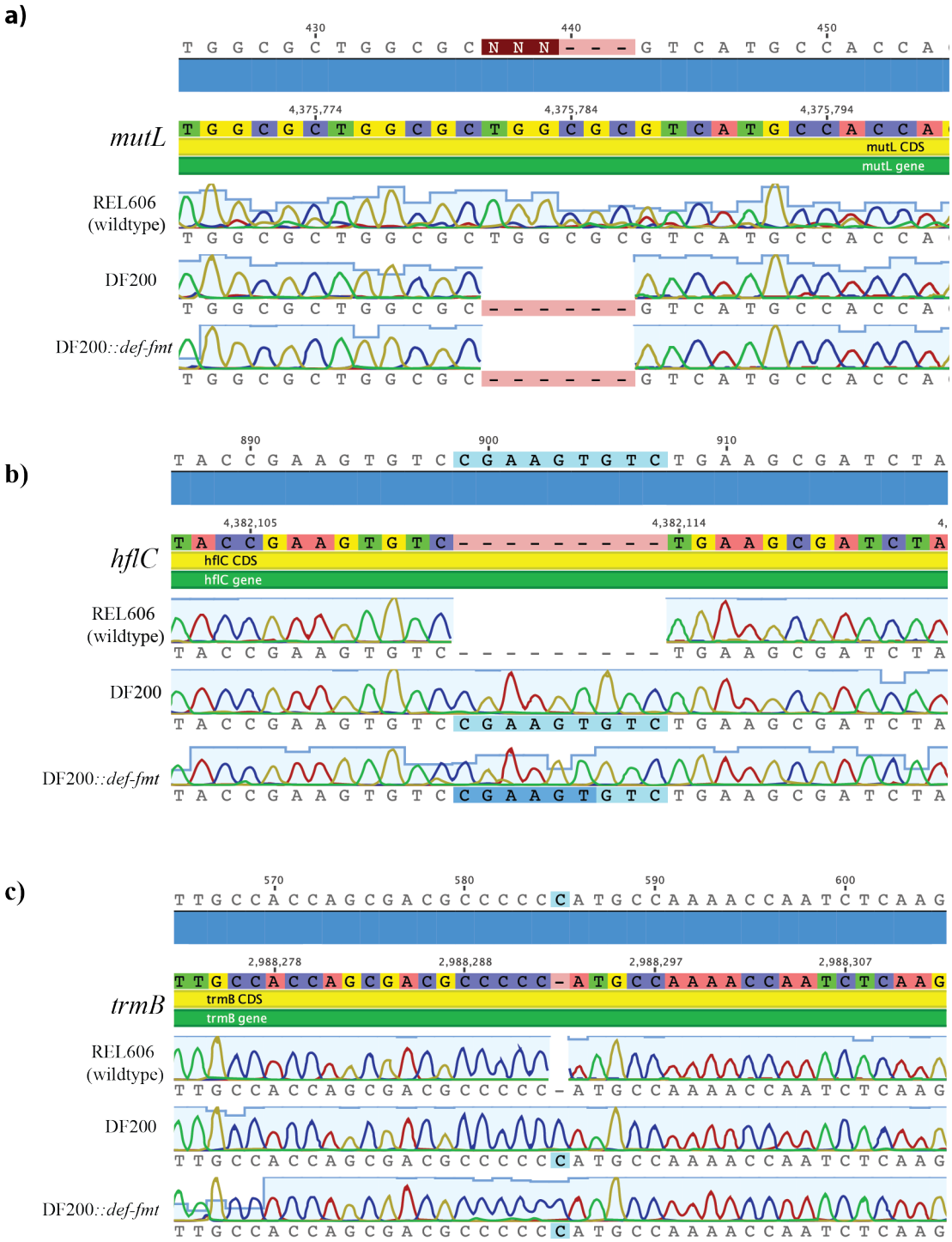


Figure 5. **Sequencing was performed in three genes known to differ between the REL606 wildtype and DF200, in order to confirm that a knock-in had been performed in the correct strain. a)** Sequencing of the *mutL* gene shows a 6-base pair deletion in both DF200 and DF200::*def-fmt*, but not in the REL606 wildtype. **b)** The *mutL* gene shows a 9-base pair insertion in both DF200 and DF200::*def-fmt*, which is not present in the REL606 wildtype. **c)** Sequencing of *trmB* reveals a single base pair insertion in both DF200 and DF200::def-fmt, but not in REL606.

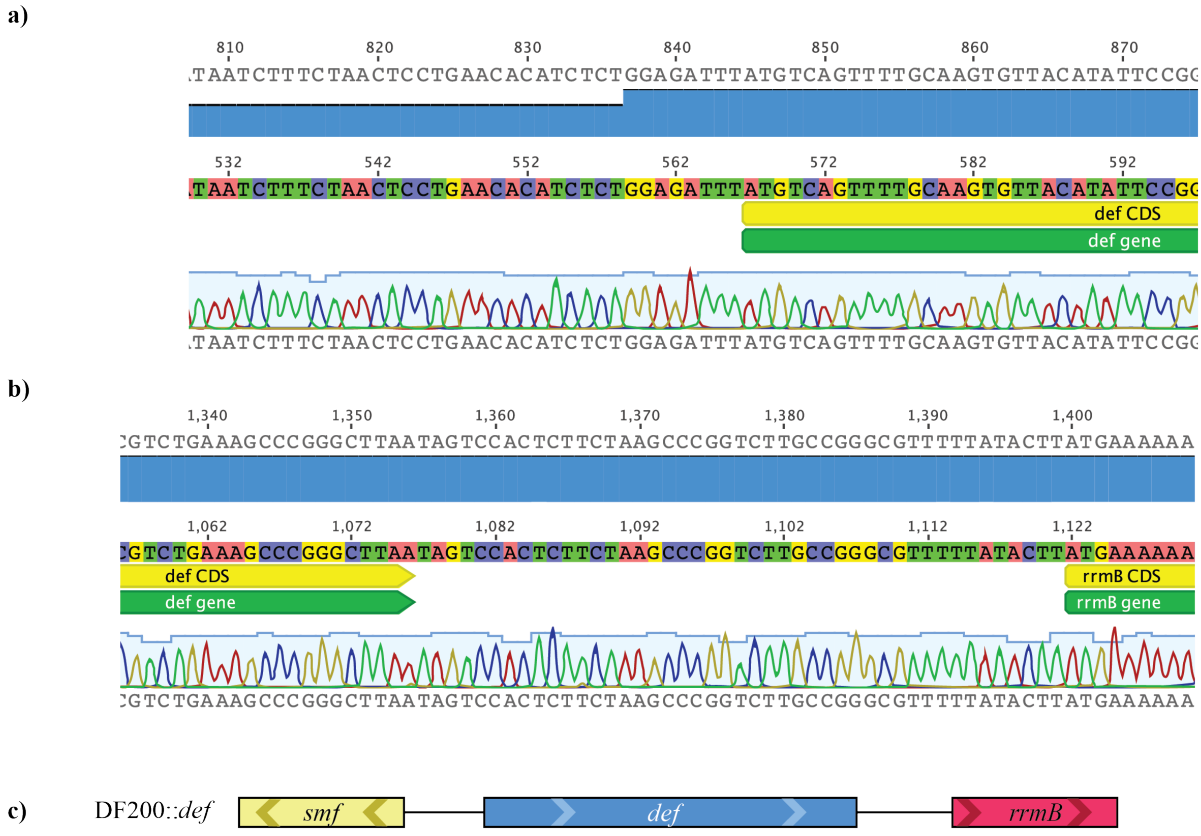


Figure 6. **Sequencing reveals correct introduction of *def* into the DF200 genome**. Sequencing was performed using primers upstream and downstream of the gene to ensure that *def* had been introduced into the correct place on the genome. **a)** Sequencing upstream of *def* reveals that the knock-in has been performed into the correct part of the genome. **b)** Sequencing was also performed downstream of the *def* gene pair. This revealed that only the *def* gene had been inserted into the genome, and that *fmt* was not present. **c)** This knock-in yielded the DF200::*def* line, a control line to be used for later experiments.

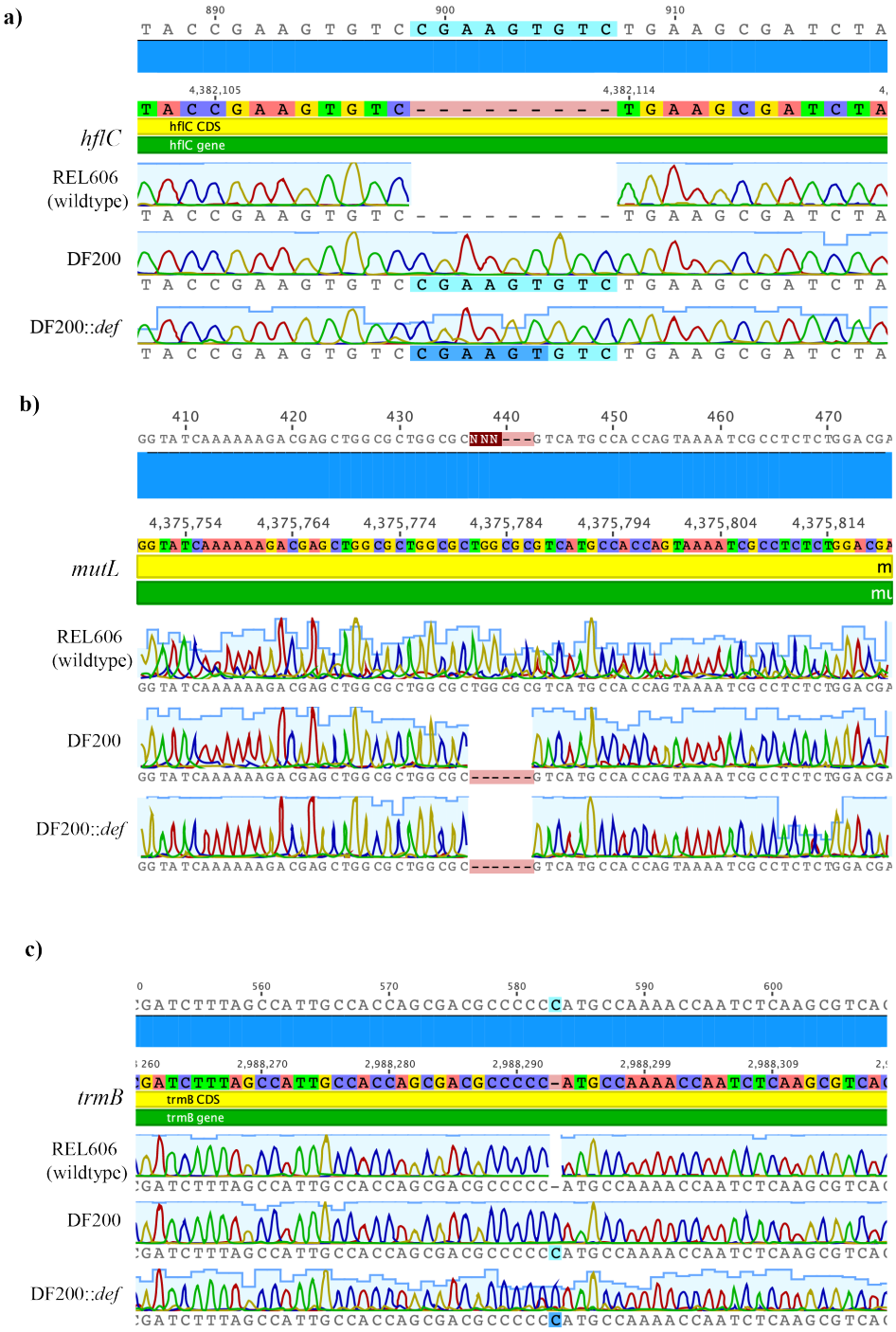


Figure 7. **Sequencing was performed for three genes known to differ between the REL606 wildtype and DF200, in order to confirm that a knock-in had been performed in the correct strain. a)** Sequencing of the *hflC* gene shows a 9-base pair insertion in both DF200 and DF200::*def*, but not in the REL606 wildtype. **b)** The *mutL* gene shows a 6-base pair deletion in both DF200 and DF200::*def*, which is not present in the REL606 wildtype. **c)** Sequencing of *trmB* reveals a single base pair insertion in both DF200 and DF200::*def*, but not in REL606.

Production of def-fmt transcripts confirmed by RT-PCR

To test whether knock-ins of the *def-fmt* and *def* genes led to gene expression, RT-PCR was conducted to elucidate whether transcription was occurring in the reintroduced regions. For the *def-fmt* knock-in, two primers were used that were internal to the *def-fmt* region. These primers would produce a single band ~1006bp in size if expression is occurring. This confirmed that a transcript was being produced from the *def-fmt* region (Figure 8). Furthermore, the single *def* knock-in was confirmed to be producing a transcript from *def*, but not from *fmt*, using two primers internal to the *def* region, and two primers internal to the *fmt* region. A band ~319bp in size would be expected from the *def* internal primers, while the *fmt* internal primers would be expected to have no band. This data reaffirms that a correct knock-in had been obtained (Figure 9).

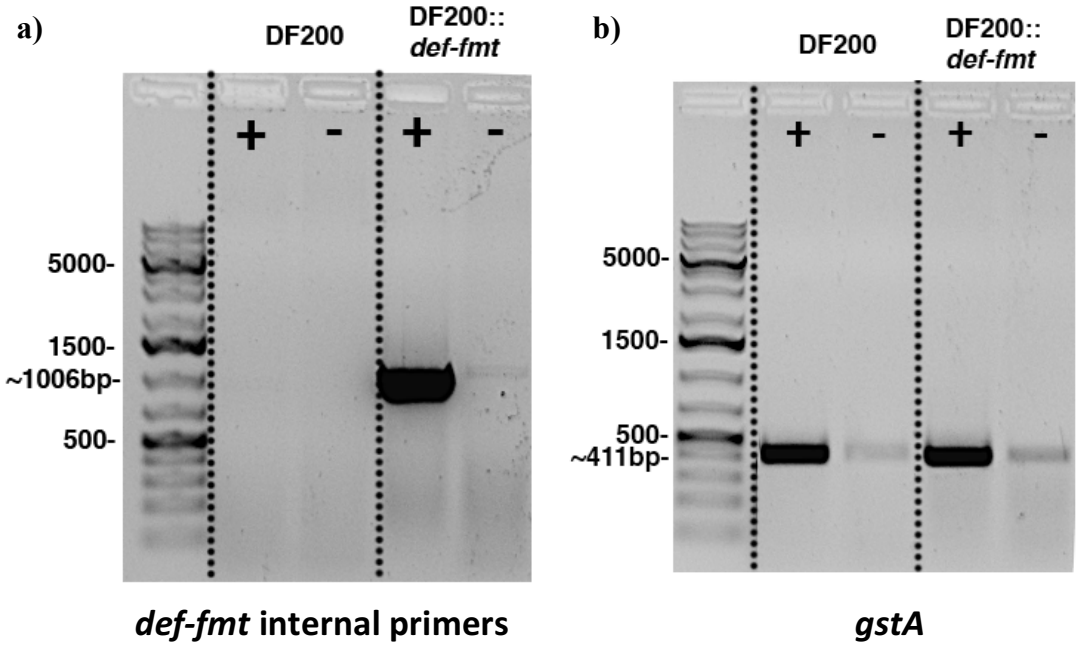


Figure 8. **RT-PCR reveals production of *def-fmt* transcripts in DF200: :*def-fmt*.** RNA was extracted from mid-late log phase cultures. Positive wells indicate a reaction subjected to a cDNA synthesis step, while the negative wells did not (refer to methods). **a)** RT-PCR performed using internal *def-fmt* primers reveals the production of *def-fmt* transcripts. **b)** primers for *gstA*, a housekeeping gene, were used as a control for RT-PCR.

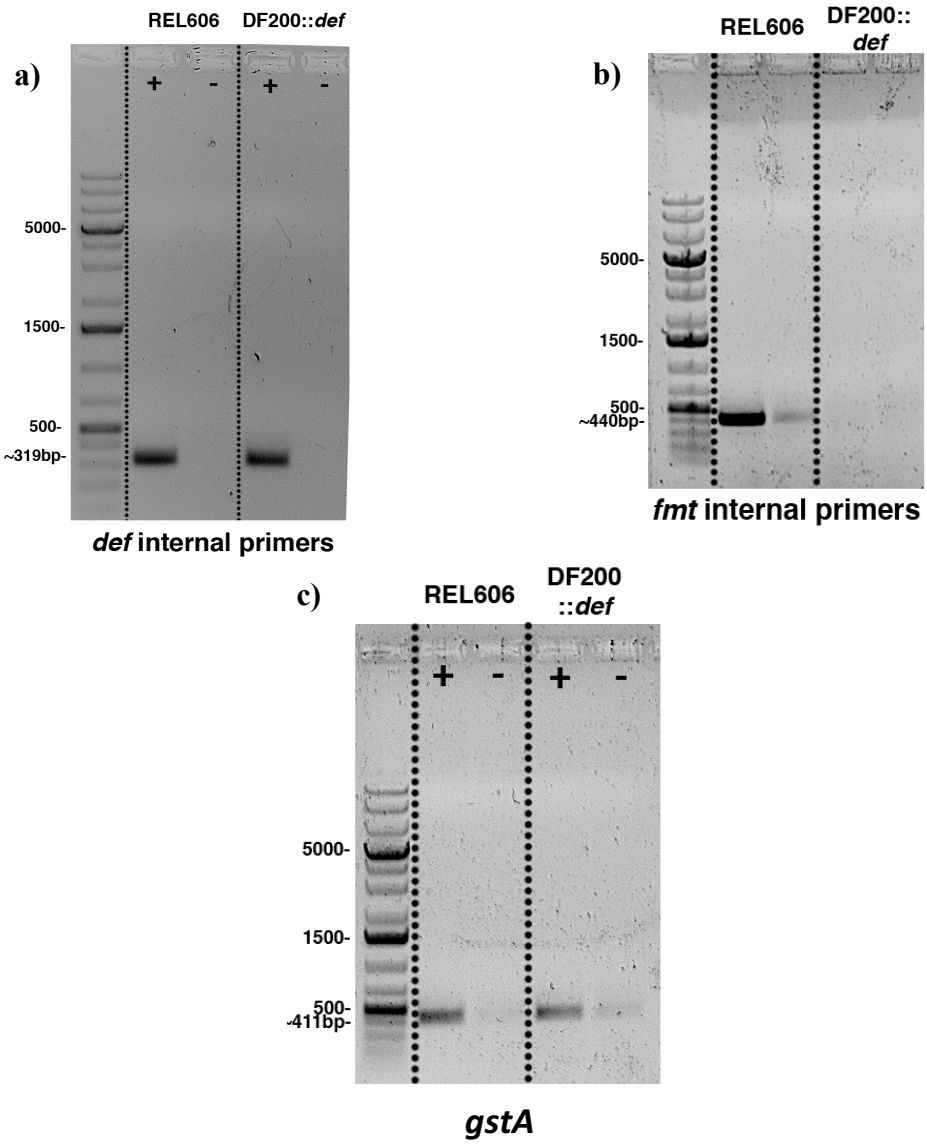


Figure 9. **DF200::*def* produces *def* transcripts, but not *fmt* transcripts**. **a)** RT-PCR performed using internal *def* primers reveals the production of *def* transcripts. Positive wells indicate a reaction subjected to a cDNA synthesis step, while the negative wells did not (refer to methods). **b)** Primers internal for *fmt* however, reveal no transcript produced in the *def* knock-in, but a transcript in REL606 wildtype. **c)** Primers for *gstA*, a housekeeping gene, were used as a control for the RT-PCR.

Northern Blot

In order to investigate whether the Met-tRNAi is being formylated, northern blotting was performed, using probes specific for the initiating tRNA. Gels of a smaller length have produced faint bands (Figure 10), with the use of DEPC-treated reagents and TBE buffer. However, these smaller gels do not provide enough resolution to separate formylated, unformylated and uncharged tRNAi molecules. Longer gels have been unsuccessful, with no bands observed thus far. We have attempted to troubleshoot this gel by altering hybridisation times and temperatures, as well as by increasing RNA concentration, but these measures have not worked. Further work is needed in order to establish this technique for a gel of a larger length, so that we may see the required separation.

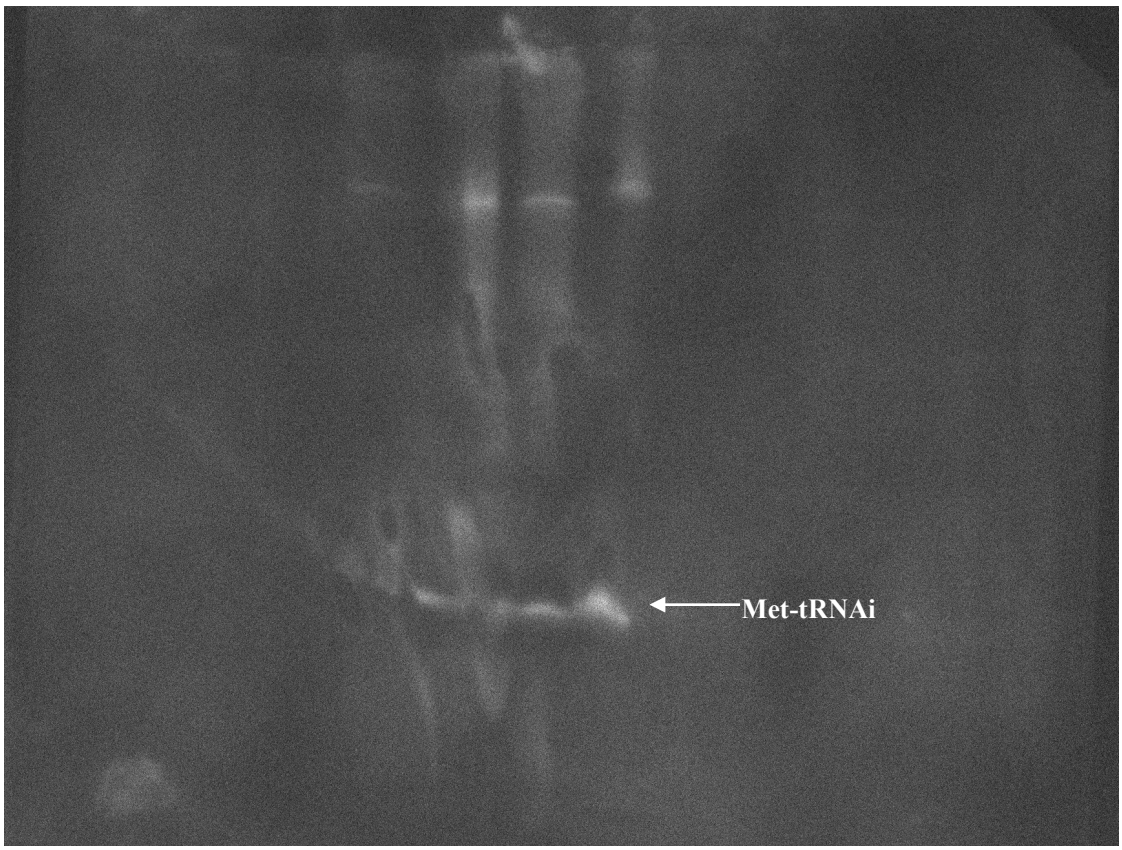


Figure 10. **Northern blotting was used in order to confirm that Met-tRNAi is formylated.** Gels at a smaller scale have shown faint bands. However, smaller gels do not provide enough resolution to differentiate between formylated and unformylated Met-tRNAi.

Knock-in ratios of def-fmt consistent with toxin-antitoxin systsem

Upon inducing recombination in the *def-fmt* knock-in, a surprising result was obtained. In theory, when recombination is induced, half of the cells should possess the original kanamycin-resistance cassette, while the other half should have the desired knock-in sequence (see methods). However, all of the cells possessed the *def-fmt* knock-in sequence when screened. This result could be due to toxin-antitoxin activity of the *def-fmt* gene pair. To investigate this, trimethoprim was used to prevent the activity of *def-fmt* having an impact on knock-in ratios. Trimethoprim is an antibiotic that prevents the reduction of folic acid, resulting in translation proceeding without the addition of the formyl group to the initiating methionine (Harvey 1973). Consequently, if an addiction effect is skewing the ratio of *def-fmt* knock-in cells, treatment with trimethoprim should eliminate this skew. The recombination step was repeated in two parallel DF200+pST76-A::*def-fmt*+pBAD-I-SceI cultures, with one of the cultures being treated with trimethoprim. Resulting colonies were then screened, and the ratio of the two sequences was analysed by digesting the PCR product, and counting the ratio of observed knock-ins to knock-outs. Any colonies which possessed the original kanamycin-resistance cassette would present multiple bands, while the *def-fmt* knock-in sequence would remain undigested. The ratios of knock-ins was also measured in the single *def* knock-in, as it also shows that if any unusual ratio is observed, that it is the result of formylation (Table 3).

Table 3. Knock-in ratios observed upon induction of recombination with trimethoprim

Strain	Number of *def-fmt/def* knock-ins (uncut)	Kanamycin resistance (cut)	Percentage of knock-ins
DF200+pST76-A::*def-fmt*+pBAD-I-SceI with trimethoprim	18	17	51.4%
DF200+pST76-A::*def-fmt*+pBAD-I-SceI without trimethoprim	36	2	94.7%
DF200+pST76-A::*def*	41	26	61.2%

Discussion

The use of formylation in bacterial translation initiation is extraordinarily well conserved, and is seen in both bacteria and eukaryotic organelles of bacterial origin. This modification appears to be important for bacterial translation, and severe impacts on cell growth are observed upon its loss (Chang *et al*. 1989; Helgren *et al*. 2016), yet formylation is not utilised in either archaea or eukaryotes. No definitive role has been defined for formylation, and experiments have shown it to be dispensable in bacterial translation (Samuel *et al*. 1970; Samuel & Rabinowitz 1974; Guillon *et al*. 1992, 1996; Newton *et al*. 1999a; Li *et al*. 2000). Because of this, it is an interesting pathway to study. It is so ubiquitously conserved across bacterial translation, yet it does not appear to be essential for translation initiation to proceed successfully. Past experiments have shown that bacteria will adapt to translation without formylation after 1,500 generations to possess growth rates that are indistinguishable from wildtype (Catchpole 2014). We then tested the model that formylation evolved as a toxin-antitoxin (TA) system, addicting cells to its presence through post-segregational killing (PSK). This revealed that along with possessing many characteristics of known TA systems, the *def-fmt* gene pair was also able to elicit a PSK phenotype (see Chapter 1), supporting our hypothesis. This aids in understanding how this system evolved, not necessarily through providing any functional benefit to

bacteria, but instead making it beneficial to not be lost through addiction to its presence.

In this thesis, I have investigated the initial introduction and adaptation to formylation. To do this, I have performed a knock-in of the *def-fmt* gene pair back into a line that has evolved in the absence of these genes for 1,500 generations. This provides a good model for investigating how formylation may have originally evolved to be so broadly conserved across all of the bacteria, yet not present in the other two domains of life. If *def-fmt* is acting as a TA pair, we might expect to see an addiction to *def-fmt* re-established.

I found that *def-fmt* has reasserted its essentiality in cells, immediately after being reintroduced into the chromosome. Specifically, the knock-in recombination step resulted in solely *def-fmt* knock-ins, instead of a 50:50 ratio of knock-ins to the original kanamycin resistance cassette, as was expected to be seen. I interpret this to be consistent with what would be expected if the *def-fmt* gene pair has toxin-antitoxin properties. If this gene pair were acting as a TA system, upon induction of recombination, the proposed toxin would be able to exert its toxic effects once again, preventing the original kanamycin-resistance genotype from being seen (Figure 11). This is due to how TA systems manifest themselves, through causing an addiction to their presence. The toxin is more stable than the antitoxin, meaning that if the gene pair is interrupted, we would expect to see growth inhibition or cell death. In our case, we did not observe any kanamycin-resistant

cells. This is consistent with our TA model (Figure 11). The reason we do not observe these cells is due to disruption of *def-fmt*, and subsequent toxicity to any cells that lose the gene pair due to *def*, our antitoxin, breaking down faster than the toxin. So, in brief, all those cells which received the kanamycin-resistance cassette during recombination were susceptible to the effects of the toxin, *fmt*.

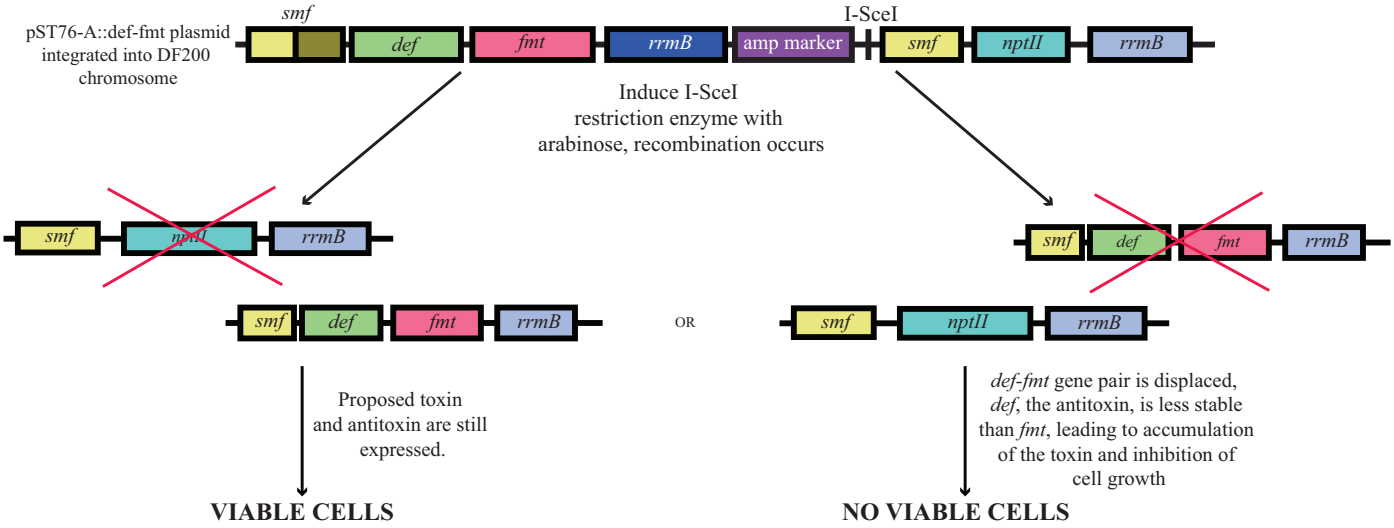


Figure 11. **Model for TA activity of *def-fmt* upon induction of recombination**. When reintroducing the *def-fmt* gene pair back onto the *E.coli* chromosome, we have used a scarless genome engineering method that should result in a 50:50 ratio of those containing the *def-fmt* gene pair, to those still possessing the kanamycin-resistance cassette (*nptII*) of DF200. This ratio has not been observed. Instead we have seen 100:0 *def-fmt*:*nptII*. We propose that we are not observing such a ratio due to toxin-antitoxin activity exerted by the *def-fmt* gene pair. Upon induction of recombination with arabinose, half of the recombinants receive the *def-fmt* gene pair, while the other half receive the kanamycin-resistance cassette. Those that receive the *def-fmt* gene pair are viable, as both the toxin and antitoxin are expressed. However, in those lines that received the kanamycin cassette, the *def-fmt* gene pair has been displaced. This leads to *def*, the antitoxin, degrading faster than the toxin (*fmt)* due to its less stable nature. As a result of this, the cells are now exposed to the toxicity of *fmt*, so that no viable cells are produced. The results we obtained are consistent with this model.

Further to this, treating the knock-ins with trimethoprim, an antibiotic that stops the use of formylation (Harvey 1973), results in seeing the original 50:50 ratio that was expected. This provides further evidence that *fmt* is acting as a toxin capable of killing recombinants that do not receive an intact *def-fmt* gene pair. These results are consistent with previous work by Ryan Catchpole (2014), where the *def-fmt* gene pair was found to act like a post-segregational killing system (PSK) (see Chapter 1). In this experiment, conditions in which post-segregational killing might occur were induced by placing the *def-fmt* gene pair on a temperature sensitive plasmid. In this way, it is possible to force the plasmid out of the cell by growing it at a temperature non-permissible to plasmid replication, replicating a situation where PSK might be induced. When the growth rate of a *def-fmt* containing plasmid was compared to the growth rate of a known PSK system PaeR7, and an empty vector, *def-fmt* was found to elucidate the same response as the known PSK system, where cell death was observed upon displacement of the gene pair. This provides strong evidence for *def-fmt* acting as a toxin-antitoxin pair. The results I have obtained upon knocking *def-fmt* back into the DF200 lines, are consistent with what would be expected to be seen upon disruption of a TA pair, and results we have previously seen (Catchpole 2014), making it a plausible explanation for the unusual knock-in ratios observed.

These data shed some light into how *def-fmt* might have originally evolved, and have spread so far. Many PSK systems are able to spread due to their addictive nature, and the fact that they are present on mobile genetic elements. If *def-fmt*

had originally existed on a mobile genetic element, then it is plausible that it could have acted as a post-segregational killing system, spreading through its ability to addict cells to its presence. Previous experiments have found that genetic addiction can be of benefit in the presence of a competitor genetic element, where it can prevent further plasmids invading (Cooper *et al*. 2010), although this is very dependent on spatial structure of the cells (Mochizuki et al. 2006). An early plasmid-bourne *def-fmt* could have acted as a protection against competitor mobile elements. The effects of its introduction were tolerated, and its addictive nature meant that it was able to become widespread. It is therefore likely that *def-fmt* was introduced into a very early ancestor of bacteria, and because of PSK, bacteria have evolved over time to accommodate its presence. TA systems allow genetic elements to persist, even if they pose some negative impact on fitness. This was seen when *def-fmt* was reintroduced into lines that had evolved without its presence, and will be discussed in Chapter 3. An immediate addiction to the presence of *def-fmt* appears to have been reasserted, showing how such systems become rapidly established when introduced.

Further experiments in this thesis aim to investigate the evolution of *def-fmt,* utilising experimental evolution to investigate how our lines adapt to *def-fmt* presence. These results help to address how *def-fmt* became so broadly conserved in bacterial translation initiation.

Chapter 3
Evolution Experiments

Introduction

In bacteria, as well as eukaryotic organelles, the methionyl-initiator tRNA is formylated and subsequently deformylated in a process that is conserved across bacteria (Giglione *et al*. 2000). Formylation, however, is not used by archaea or elsewhere in eukaryotes, making it an intriguing pathway for studying the evolutionary origins of bacteria. It is not an essential process, with knock-outs of the gene pair producing viable cells with a significant decrease in growth rate, although varying effects are seen across different bacterial species (as discussed in Chapter 1) (Guillon *et al*. 1992). We have previously used evolution experiments in *Escherichia coli* to investigate the evolution of this highly conserved process. Experimental evolution has been successfully utilised in a range of bacterial species to investigate evolutionary response to a number of situations, including limiting nutrition (Lenski *et al.* 1991), increases of temperature (Cullum *et al.* 2001), and the evolution of antibiotic resistance (Gullberg *et al.* 2011). In this work, the use of evolution experiments would provide us with an ideal model system for studying the early evolution of formylmethionine use, and may provide answers as to how it became so universally conserved. We previously utilised such evolution experiments to investigate the capacity of *E. coli* REL606 to survive without the *def-fmt* gene

pair (Catchpole 2014). Despite an initial detriment to growth rate, these lines were found to have a growth rate that was indistinguishable from the wildtype after 1,500 generations of culturing. Whole genome sequencing of these lines revealed a large array of parallel and compensatory mutations in response to the loss of *def-fmt.* To further investigate how formylation initially evolved, we have now knocked the *def-fmt* gene pair back into these evolved lines and have allowed them to evolve for a further 3,000 generations.

We have found that the reintroduction of the *def-fmt* gene pair has a minimal impact on cell fitness, even after these lines have evolved devoid of the gene pair. Further to this, 3,000 generations of long-term culturing with the gene pair reintroduced has no significant impact on the growth rate of these cells. Our results indicate that an addiction to *def-fmt* has been reasserted. Intriguingly also, is the fact that this addiction appears to be immediate upon reintroduction of the *def-fmt* gene pair, despite these lines having evolved in its absence for 1,500 generations. This addiction is something that is characteristic of known toxin-antitoxin systems, providing evidence that the use of formylmethionine in bacterial translation initiation could have evolved to be so ubiquitous through such a system.

Methods

Strains and media

All strains used in this chapter, have already been previously defined (see Chapter 2). DF200, DF200::*def-fmt*, DF200::*def* and REL606 strains are used as specified. All strains were grown at 37°C.

Media used for these experiments was Davis Minimal (DM0) (Difco), supplemented with 200mg/L dextrose (DM200) or 2000 mg/L (DM2000) and 2 mg/L thiamine or LB (Oxoid) as specified. Solid media was made by adding bacteriological agar to a concentration of 1.5% w/v (Oxoid). The following antibiotics were also used as specified at the following concentrations: streptomycin, 100 μg/ml; ampicillin, 100μg/mL; actinonin, 5 μg/mL, 10 μg/mL, 20 μg/mL, 50 μg/mL 100 μg/mL and 200 μg/mL; chloramphenicol, 20μg/mL; kanamycin, 20μg/mL.

Serial culturing

DF200::*def-fmt* and DF200::*def* cultures were plated on DM200 agar with streptomycin and left to grow overnight at 37°C. Single colonies were selected and grown to stationary phase in DM200 media supplemented with streptomycin. Each culture was then diluted 1:100 in DM200 with streptomycin, with 11

replicates of each line. A media control was also run alongside the 11 replicates in order to ensure media sterility. These 11 replicates were serially cultured in this way for 100 days, and after each transfer, glycerol stocks were made by adding 50% glycerol to a final concentration of 15% and storing at -80°C. Once per week, cultures were tested for contamination. This was done by plating cultures on tetrazolium arabinose agar (Levin *et al.* 1977; Lenski *et al.* 1991), as well as testing sensitivity to T4 and T5 phage. Single colonies were also PCR screened with the def-808bp and fmt+678bp primers listed below:

def-808bp:

5′ –GCTTCCACCACCAGTACACC– 3′

fmt+678bp:

5′ –CCAGACGTACAGCATCAGGG– 3′

Measurement of growth rates

Cultures of DF200, REL606 and varying DF200::*def-fmt* lines were grown overnight at 37°C in DM200 media with streptomycin. The DF200 cultures were also supplemented with kanamycin. A 1:100 dilution was made of these cultures, and were left to grow for 16 hours with shaking at 37°C with OD_{595} monitored by the FLUOstar OMEGA (BMG Labtech). Growth rates were then determined

by calculating the minimum doubling time over a 30-minute time period. Three replicates were performed for each of the 11 lines for each ten-day period.

Actinonin Minimal Inhibitory Concentrations (MIC)

Three replicate overnight cultures of DF200::*def-fmt* and DF200::*def* at both day 0 and day 100, and REL606 were grown at 37°C. 1:100 dilutions of these cultures were made in DM200 supplemented with the following concentrations of actinonin: 0 μg/mL, 5 μg/mL, 10 μg/mL, 20 μg/mL, 50 μg/mL, 100 μg/mL and 200 μg/mL. The DF200 lines were also supplemented with kanamycin as previously specified. The growth of these cultures was then monitored for 24 hours at 37°C with a FLUOstar OMEGA plate reader (BMG Labtech). This yielded three biological replicates for each line at each concentration of actinonin.

Knock-outs of def-fmt gene pair

To achieve a knock-out of the *def-fmt* gene pair in DF200::*def-fmt* lines, a gene gorging protocol was used (Herring *et al.* 2003). Cultures of REL606 +pGEM:I-SceI-Def-Kan-Fmt and DPWC+pACBSR were obtained from Ryan Catchpole (Catchpole 2014). The first plasmid contains a *nptII* cassette, conferring kanamycin resistance from the plasmid pKD4 (Genbank: AY048743.1) (Datsenko & Wanner 2000), as well a I-SceI cut site. An overnight culture of this line was grown in LB supplemented with chloramphenicol, ampicillin,

kanamycin, streptomycin, and 1% glucose at 37°C. DPWC+pACBSR was also grown overnight in LB supplemented with chloramphenicol and 1% glucose to suppress lambda red activity. This second plasmid expresses λ recombinase and I-SceI, making it possible to replace the *def-fmt* gene pair with *nptII*, a kanamycin resistance cassette. The plasmids from both aforementioned cultures were then isolated using the Purelink® Quick Plasmid Miniprep Kit (Invitrogen).

Overnight cultures of both day 0 and day 80 DF200::*def-fmt* cultures were grown overnight at 37°C in DM200 supplemented with streptomycin. Day cultures of these cells were then grown to mid-late log phase at 37°C, and were used for transformation using calcium chloride and heatshock (Sambrook *et al.* 1989) with the pACBSR plasmid. Transformants were selected by plating on DM2000 agar with streptomycin, ampicillin and chloramphicol, supplemented with 1% glucose. Observed colonies were screened by PCR using the KAPA2G Robust HostStart ReadyMix PCR Kit (KAPA Biosystems) and the following primers:

pBAD_F:

5′ –ATGCCATAGCATTTTTATCC– 3′

pBAD_R:

5′ –GATTTAATCTGTATCAGG– 3′

Colonies yielding the correct PCR product were then used for transformation with the pGEM:I-SceI-*def*-Kan-*fmt* plasmid, and were plated

on DM2000 agar containing streptomycin, ampicillin, chloramphenicol and kanamycin at 37°C for 16 hours. Colonies were screened by PCR using the M13 primers:

M13_F:

5′ –GTAAAACGACGGCCAGT– 3′

M13:R:

5′ –GCGGATAACAATTTCACACAGG– 3′

Recombination was then induced in the correct colonies by growing the cultures in DM0 supplemented with 1% arabinose, streptomycin and kanamycin for 24 hours at 37°C. 10^{-3}, 10^{-4}, 10^{-5} and 10^{-6} dilutions of this culture were grown on DM2000 agar with 1% arabinose, streptomycin and kanamycin. Observed colonies were then screened using the def-808bp and fmt+678bp primers below, and were digested with the SphI enzyme (New England Biolabs) for 2 hours at 37°C.

def-808bp:

5′ –GCTTCCACCACCAGTACACC– 3′

fmt+678bp:

5′ –CCAGACGTACAGCATCAGGG– 3′

Results

Growth of DF200 is marginally decreased by reintroduction of def-fmt

Previous work showed that the growth rate of wildtype *E. coli* was severely impacted upon the knock-out of the def-fmt gene pair, although these lines evolved to possess a growth rate indistinguishable from the original wildtype. In order to investigate whether an effect was seen as the result of knocking the gene pair back into the evolved DF200, the growth of DF200 and DF200:*def-fmt* was monitored over 17 hours using a FLUOStar OMEGA plate reader. While it appeared that there was a small impact on growth rate (calculated as a minimum doubling time over a 30 minute period), upon reintroduction of the *def-fmt* gene pair, this was not significant (Figure 1) (student's t-test, p=0.397).

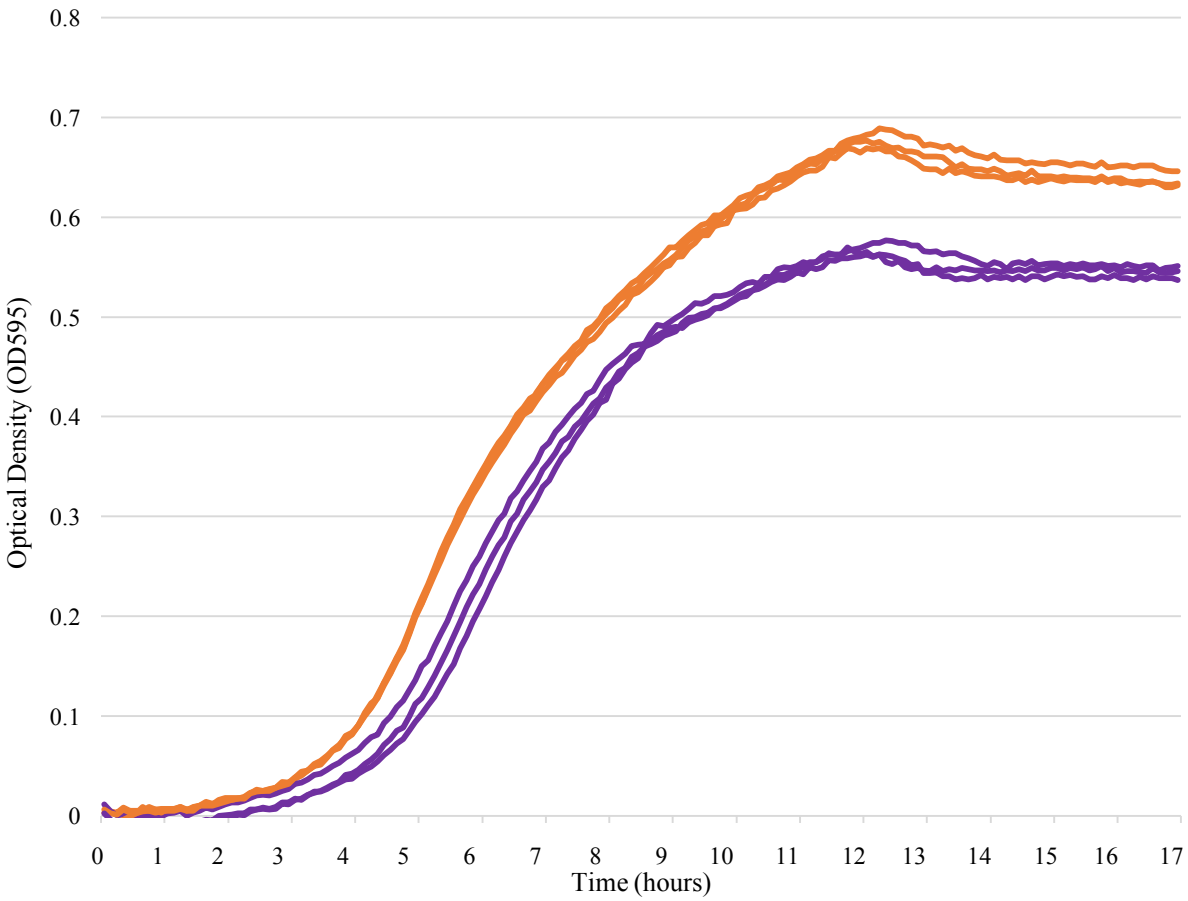


Figure 1. **Growth rate of DF200 is minimally impacted by the reintroduction of the *def-fmt* gene pair.** Three DF200 clones (orange) and three DF200::*def-fmt* clones (purple) were grown for 17 hours at 37°C. The optical density of these cultures was measured over this time using a FLUOStar OMEGA plate reader. Growth rate among the three replicates of each line is nearly identical. While the growth rate of DF200::*def-fmt* appears to have decreased in response to the reintroduction of the *def-fmt* gene pair, the difference is not significant (t-test, p=0.397).

The growth rate of DF200::def-fmt remains unchanged after 3,000 generations of long-term evolution

The growth rates of the evolving lines were monitored throughout the duration of the long-term evolution experiment. The growth rate was measured at ten-day intervals for 100 days during the culturing, and was compared to the original DF200 ancestor. Over the course of culturing, it can be seen that there was minimal change to the overall growth rate of the cultures (Figure 2). Further to this, a Wilcoxon rank-sum test was used to assess whether there was a significant difference in the growth between the evolving line and its ancestor. This test provided a non-significant result ($p=0.65$), indicating that there is no significant difference between the lines.

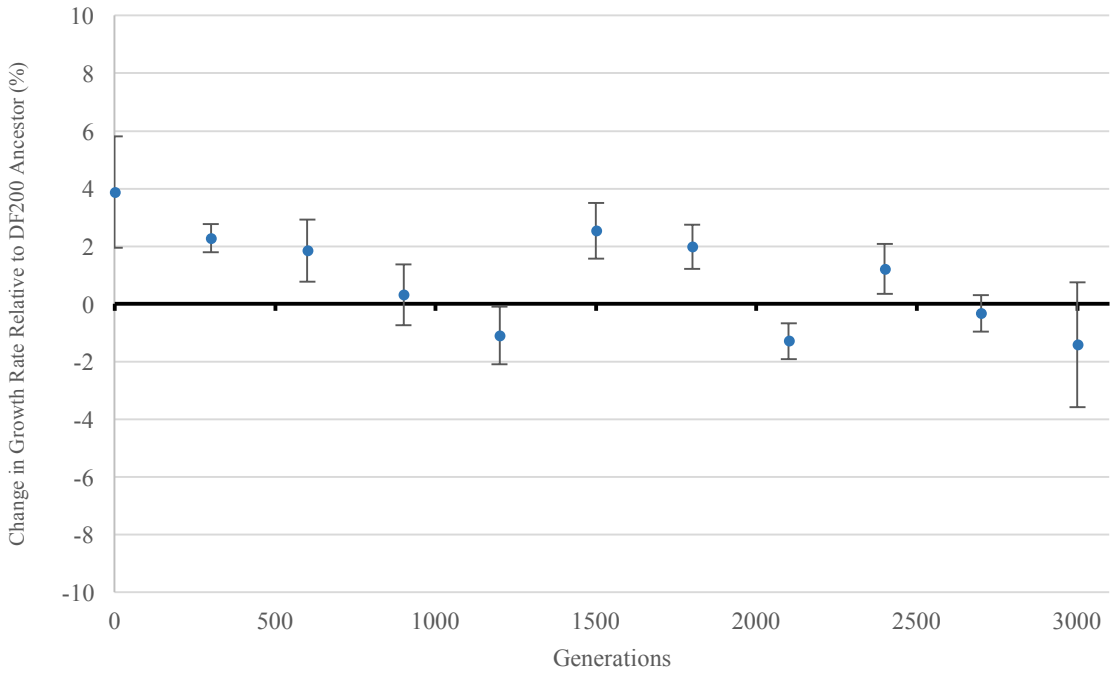


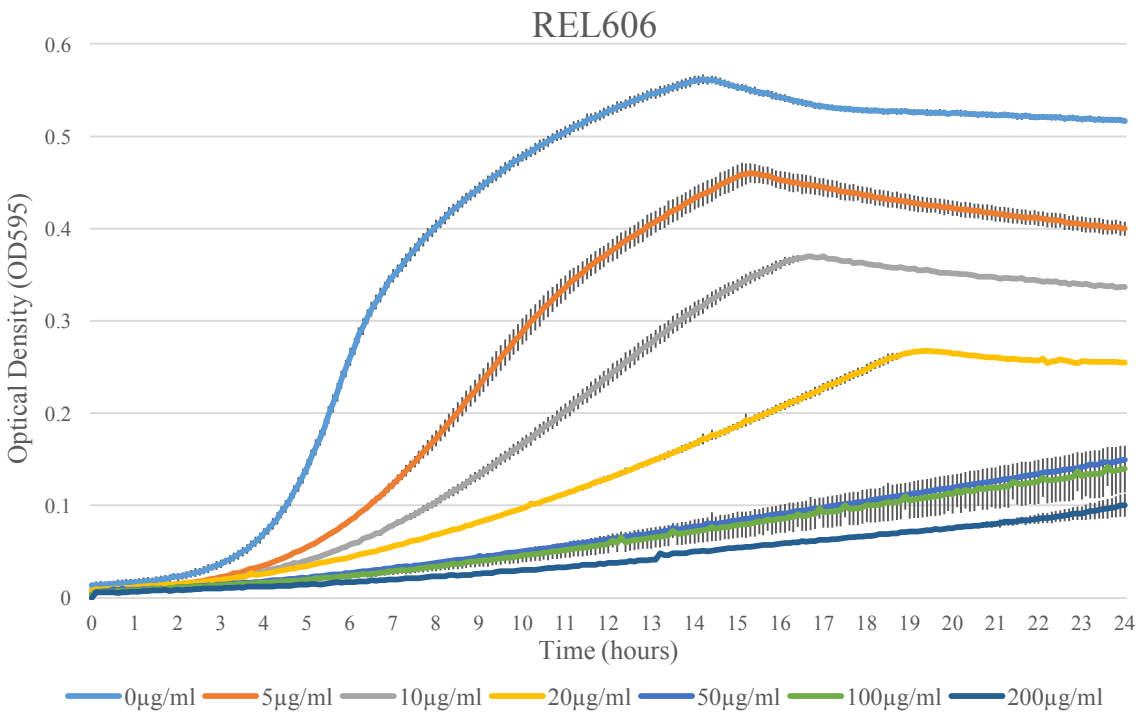
Figure 2. **The growth rate of DF200::*def-fmt* evolved lines did not change significantly over 100 days of culturing.** The growth of the lines was monitored at ten-day intervals, and doubling times for each line were calculated as the minimum doubling time over a thirty-minute period. Each point represents the average of the 11 lines ($\bar{x}$±SE). This is compared as a percentage difference to the growth rate of the DF200 ancestor, with negative values representing a decrease in doubling time, or an increase in growth rate. The calculated doubling time of DF200 is 45.6 minutes.

Lines with def-fmt reintroduced are no longer resistance to actinonin.

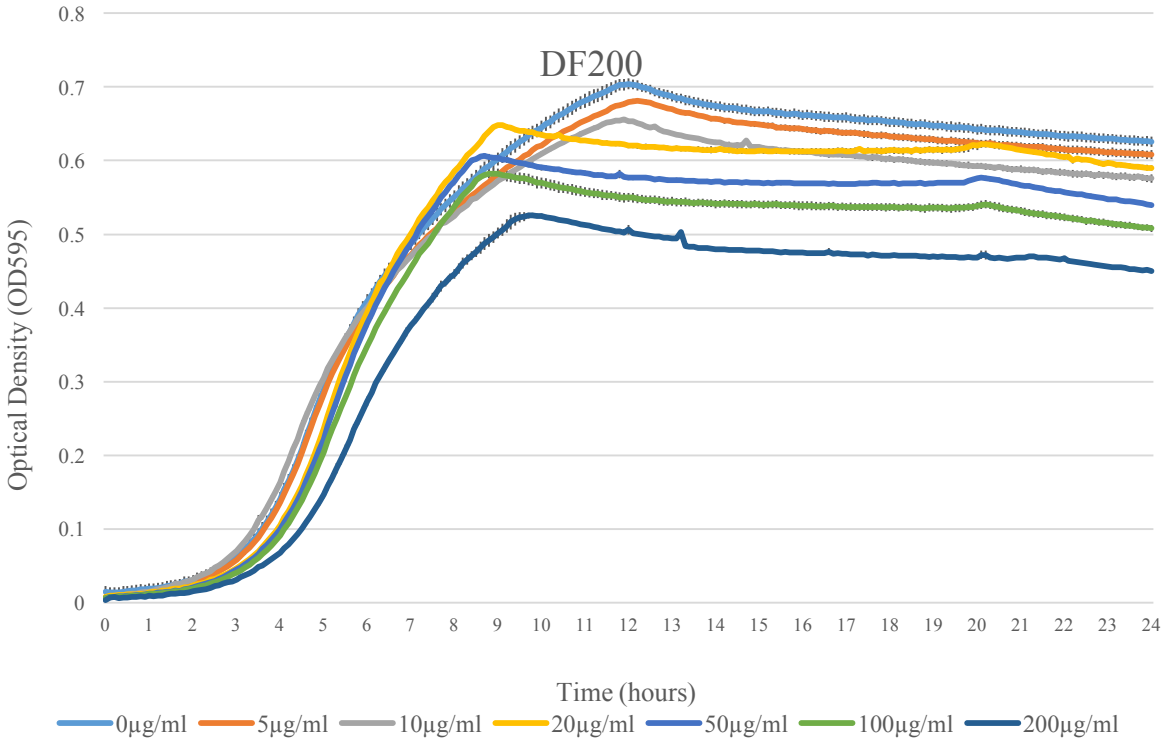
Upon knock-out of *def-fmt*, previous work has shown a resistance to actinonin, an antibiotic that targets peptide deformylase (Catchpole 2014). In order to investigate the impact of reintroduction of *def-fmt*, the lines were grown in increasing concentrations of actinonin (Figure 3). Actinonin acts by inhibiting peptide deformylase, resulting in death of cells that are formylating (Yuan *et al.* 2001). Treatment with actinonin upon reintroduction of *def-fmt* resulted in susceptibility to the antibiotic returning.

Interestingly, a single knock-in of peptide deformylase is still resistant to actinonin, even after 100 days of long-term evolution, indicating that even when actinonin itself targets peptide deformylase, its effects are not seen unless formyltransferase is also present. Some inhibition of growth is observed in the resistant lines, although this might be attributed to the use of DMSO as a solvent when dissolving the actinonin. Concentrations of DMSO used in this experiment have been previously shown to negatively impact bacterial cell growth (Markarian *et al.* 2002).

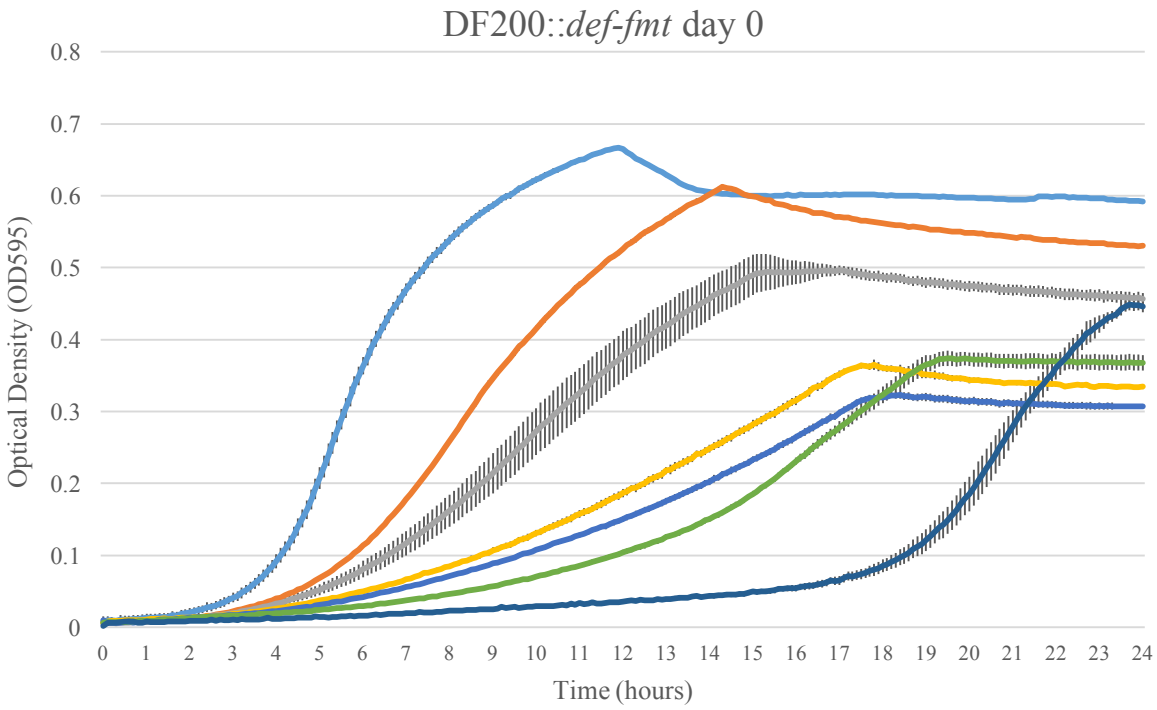
a)


REL606
Optical Density (OD595)
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)
0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml

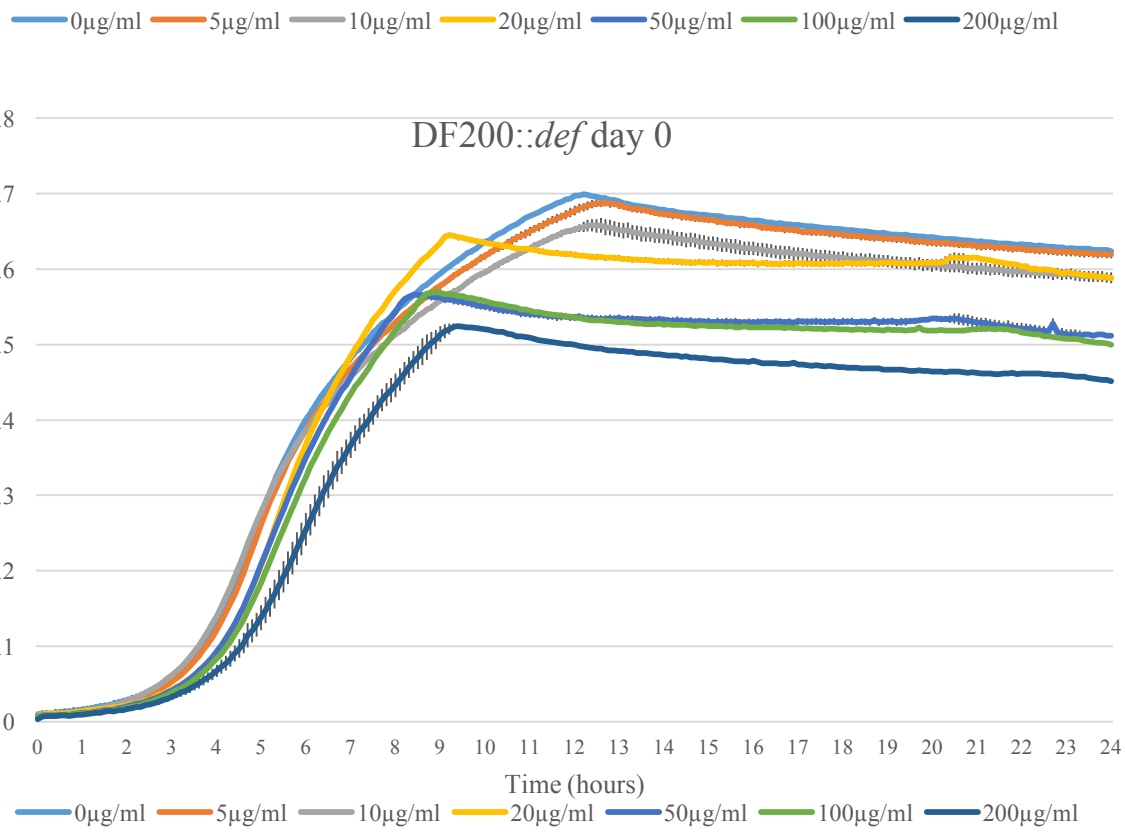
b)


DF200
Optical Density (OD595)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)
0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml

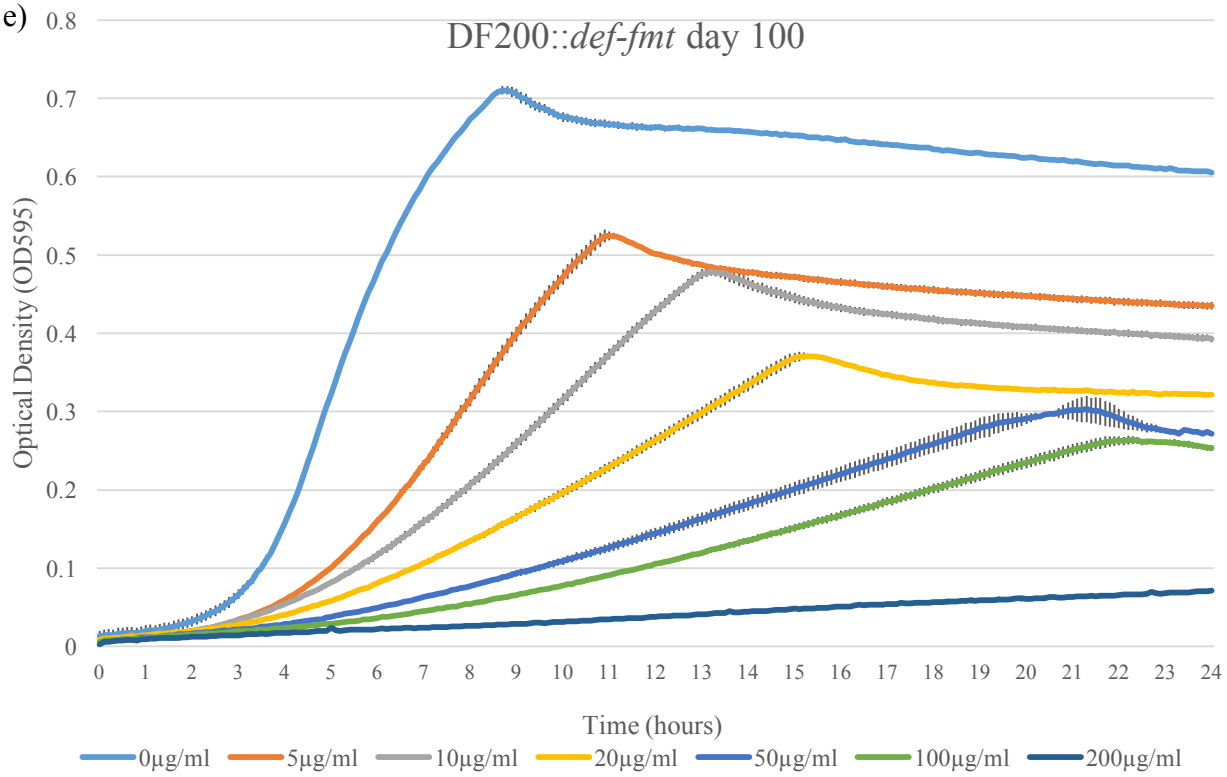
c)


DF200::def-fmt day 0
Optical Density (OD595)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)

d)


0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml
DF200::def day 0
Optical Density (OD595)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)
0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml

e)


DF200::def-fmt day 100
Optical Density (OD595)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)
0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml

f)

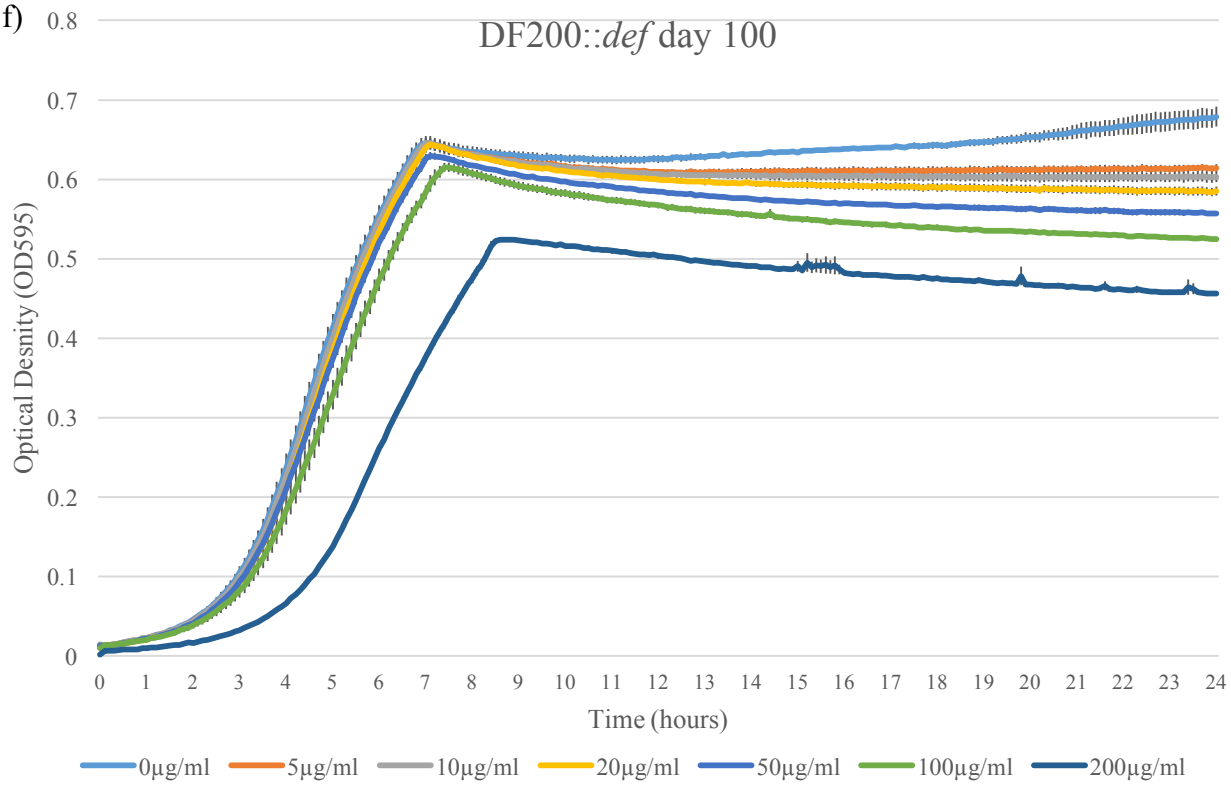

DF200::def day 100
Optical Desnity (OD595)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Time (hours)
0µg/ml
5µg/ml
10µg/ml
20µg/ml
50µg/ml
100µg/ml
200µg/ml

Figure 3. **Susceptibility to actinonin is reasserted upon reintroduction of the *def-fmt* gene pair**. In order to assess the impact of reintroduction of the *def-fmt* gene pair, cultures were grown in varying concentrations of actinonin and growth was monitored over 24 hours ($\bar{x}$±SE). **a)** Growth of REL606 (wildtype) in increasing concentrations of actinonin shows increasing susceptibility as the concentration increases. **b)** Actinonin is found to have a minimal effect on cell growth in DF200 (a *def-fmt* knock-out evolved in the gene pair's absence for 1,500 generations). **c)** Upon reintroduction of the *def-fmt* gene pair (DF200::*def-fmt* day 0), susceptibility to actinonin is observed once again. **d)** Intriguingly, reintroduction of the *def* gene alone (DF200::*def* day 0) does not result in increased susceptibility to actinonin, despite the gene being the primary target of the antibiotic. **e)** After 100 days, or 3,000 generations, of long-term culturing, susceptibility of the DF200::*def-fmt* line to actinonin is still present. **f)** A single knock-in of *def* does not appear to be any more or less susceptible to actinonin after 100 days of long-term culturing, indicating that formylation is still not occurring.

Attempts at knocking out def-fmt reveals addictive phenotype

In order to investigate the impact of *def-fmt* reintroduction, knockouts of the *def-fmt* gene pair were attempted at both day 0 and day 80 during the long-term evolution experiment. This was done through a gene gorging protocol (Herring *et al.* 2003), and utilised lambda red recombinase in order to replace the *def-fmt* gene pair with a kanamycin resistance cassette. Upon screening the potential knockout cells, all were still observed to still possess the original *def-fmt* sequence (Table 1). This indicates that no successful knockouts were obtained.

Table 1. Observed colonies upon attempted replacement of the *def-fmt* gene pair with a *nptII* cassette.

	*Successful knock-outs (*ntpII*)*	*Unsuccessful knock-outs (*def-fmt*)*	*Percentage of successful knock-outs (%)*
Day 0	0	100	0
Day 80	0	100	0

Discussion

We have previously proposed a model whereby *def-fmt* originally invaded bacteria on a selfish genetic element and persisted through toxin-antitoxin (TA) activity. These systems defy conventional beliefs of genes persisting due to an inherent benefit being conferred to cells (Mochizuki *et al.* 2006). With post-segregational killing systems (PSK), we see a situation where their introduction is not necessarily beneficial, but instead a situation where it is detrimental to lose the addiction system once it is present (Cooper & Heinemann 2000). For *def-fmt* to have evolved to be so ubiquitous in bacterial translation initiation, it might be expected that there would have been a large fitness benefit associated with its introduction. This, however was not seen in our experiments, making it likely that something beyond a fitness benefit was responsible for its presence across all bacteria. We are not observing an inherent fitness benefit upon reintroduction of *def-fmt*, what we do see is an immediate reassertion of addiction to its presence through an inability to knock these genes out, something that is consistent with TA systems. The growth rate of these bacteria was also unchanged over 3,000 generations of long-term culturing. These results would indicate that there have been minimal fitness costs associated with *def-fmt* reintroduction, despite the DF200 ancestor having evolved in its absence for 1,500 generations. Such observations are expected under our model of PSK, where we are not seeing a fitness benefit associated with *def-fmt* reintroduction.

We did, however, observe a slight decrease in overall optical density between our knock-in and ancestor lines. As these lines possessed no significant difference in growth rate, this observation could be due to a smaller cell size in our knock-in line. This could be tested in the future through CFU (colony forming units) counts. If we observe the same number of colonies between the two counts, it is likely that our difference in optical density is due to a decrease in cell size.

A capacity for post-segregational killing alone is not sufficient for complete spread and maintenance of a toxin-antitoxin system (Cooper & Heinemann 2000). Previous experiments have shown that the spread of PSK systems is heavily reliant on environmental factors. This was seen in theoretical work undertaken by Mochizuki *et al.* (2006). Mochizuki found that genetic addiction was dependent on spatial structure of the environment the cells were growing in, and that in an environment without spatial structure, it is very difficult for these systems to spread. For a PSK system to spread in such a situation, an initially high concentration of the plasmid is required (Cooper & Heinemann 2000). Further to this, it has also been established that if a plasmid relies solely on PSK to persist, that selective pressure would eventually result in its loss due to inactivation of the toxin (Kobayashi 2004). In such a case, we would expect to see ordered gene loss of *def-fmt* if its capacity as a TA system alone was causing its persistence. Instead, we observe no loss, as no knock-outs were obtained following reintroduction. Further to this, we also see no loss of *def-fmt* upon

induction of recombination in cells containing a plasmid with the *def*-fmt gene pair (see Chapter 2).

The fact that we are observing no loss of *def-fmt* would indicate that a further mechanism could be acting to preserve *def-fmt* within our lines, beyond the action of the toxin-antitoxin system. Such an observation would be consistent with a model where horizontal transfer of the *def-fmt* gene pair is occurring after genome integration, and could explain why we are not seeing loss of the *def-fmt* gene pair. Under such a model, we would expect a situation where adaptive change to *def-fmt* is occurring quicker than any ordered loss. Rapid adaptation to *def-fmt* reintroduction could make loss of the gene pair disadvantageous as it is impacting the fundamental translational machinery, explaining why we are no longer able to knock it out, especially early on in our evolution experiment. These lines have previously been evolved in the absence of *def-fmt* for 1,500 generations, and we have seen evidence of adaptive change to *def-fmt* loss, with wildtype growth rates observed (Catchpole 2014). As a result of this, we would expect to see loss of *def-fmt* more frequently in the lines adapted to the absence of formylation. By reintroducing *def-fmt*, we are mimicking the initial integration event of *def-fmt*, making it interesting that we are observing no loss of *def-fmt.* This might aid in explaining how formylation has persisted for so long and became so ubiquitous in bacterial translation, through a combination of causing addiction to its presence, as well as further spread being possible through horizontal transfer. Evolution experiments provide an ideal situation in which to

test a model of horizontal gene transfer. We have subsequently evolved these knock-ins for a further 3,000 generations with the *def-fmt* pair reintroduced, and would expect to see ordered loss of the genes in some of these lines due to the nature of toxin-antitoxin systems, where loss is expected due to inactivation of the toxin (Van Melderen & De Bast 2009). A model of horizontal gene transfer, however, would explain no loss of the gene pair over the duration of the evolution experiment. Screening of cells with PCR throughout the duration of the experiment as a form of contamination check has indicated that the *def-fmt* gene pair is still present within the genome, but does not confirm whether there have been mutational changes within the genome that inactivate *fmt*, the proposed toxin. We have further investigated the genome through the use of sequencing, and have observed no mutational changes in the *def-fmt* region. This data will be discussed further in Chapter 4.

We have previously hypothesised that *def-fmt* could have been extremely evolutionarily old in order for it to be present in all bacteria, and that it initially invaded an extremely early ancestor of bacteria (Catchpole 2014). However, we would expect to see loss of formylation in more lineages if PSK is relied on for persistence, consistent with what is knock about PSK systems. Under a model of horizontal gene transfer, it is possible that *def-fmt* is not necessarily old, but that it was extremely good at spreading across bacteria, even once the gene pair was integrated into the genome.

In our experiments we have also observed susceptibility to actinonin returning instantly upon reintroduction of the *def-fmt* gene pair. This would indicate that these lines are once again susceptible to the effects of formylation, consistent with our previous observation that addiction to *def-fmt* is immediately reasserted upon knock-in of these genes. Actinonin targets peptide deformylase, and it is the presence of formylated proteins that is lethal to the cell (Yuan *et al.* 2001). This explains why, even though actinonin targets peptide deformylase itself, in the absence of formyltransferase, we do not see any inhibitory effects.

In order to investigate changes at a genetic level in my evolved lines, and potentially elucidate more details on whether we are observing the proposed model of horizontal gene transfer within the lines, we have performed whole genome sequencing. This data will be discussed in the next chapter. It is expected that if our model is correct, we would see no instances of *def-fmt* loss as well as rapid changes in parallel in these lines. Such parallel changes are expected to be adaptation to formylation occurring, and would be faster than any ordered gene loss. It would be anticipated that these parallel changes are seen within genes associates with the translational machinery in response to formylation occurring once again.

Chapter 4
Genome Analysis

Introduction

The use of formylated methionine is ubiquitous across bacteria, despite possessing no clear function. We have previously shown it to be dispensable from bacterial translation, with *E. coli* adapting to its loss within a 1,500-generation evolution experiment. These lines possess growth rates indistinguishable from the wildtype (Catchpole 2014). Evolution experiments provide an effective system for us to now study how formylation evolved and became a ubiquitous feature of bacterial translation initiation. We are able to effectively emulate the initial introduction of *def-fmt*, and monitor the response in lines we have previously evolved in the absence of these genes.

We have previously used whole genome sequencing to study the response of *E. coli* to the loss of the *def-fmt* gene pair (Catchpole 2014). Whole genome sequencing allowed us to study at a genetic level exactly how these bacteria were responding to the loss of these genes, and identify mutations that might be important to the adaptation that we saw. A number of mutations were attributed to media adaptation, and had been seen in a previous long-term culturing experiment (Barrick *et al.* 2009). However, a number of parallel mutations were seen in genes important for the translational machinery, suggesting they occurred

in response to *def-fmt* loss. SNPs were observed upstream of *metZ,* encoding the initiating met-tRNA*,* within the promoter region in all 11 lines. While 10 of these mutations were identical, one was different. In order to investigate these mutations further, a GFP reporter system was constructed, using the promoter region of *metZ* to control GFP expression. It was found that during stationary phase, expression of GFP was increased, and that the two observed mutations conferred the same phenotypic change. It was thought that these mutations could be helping to compensate for the lower affinity of initiation factor 2 (IF2) for non-formylated methionine through increasing expression of *metZ*. However, as this was performed in a GFP reporter system on a plasmid, these results may have limitations. While the reporter assay did show an increase in GFP expression within this promoter (Catchpole 2014), it does not tell us whether this mutation in *metZ* actually results in increased fitness in our lines, meaning that it is compensating for the loss of *def-fmt*. There are a large number of interacting factors involved in translation initiation, and we have seen mutations in a number of genes involved with translation in the DF200 lines. including initiation factor 2 (IF2). This means that a mutation in *metZ* might not be resulting in a large increase in fitness in our lines. Experiments that involve knocking in the wildtype REL606 sequence of the *metZ* promoter into DF200 will allow us to analyse the impacts of this mutation. If we see a decrease in fitness in response to the introduction of the wildtype sequence, it is likely that increasing *metZ* expression is important in adapting to the absence of *def-fmt.*

Mutations in *infB*, the gene that encodes initiation factor 2 (IF2), were also observed. Three different SNPs were observed across the eleven lines, with one of these SNPs occurring in ten of the eleven evolving lines. While one of these mutations had been seen in a region previously associated with media adaption (R749L), two of these mutations were found to have occurred in the C2 domain of IF2 (P807S and Y838C), the region where fMet-tRNAi^{fMet} binds. The location of these changes means that these mutations might be important for adapting to the loss of *def-fmt* (Figure 1).

The reintroduction of *def*-fmt to a line devoid of these genes for 1,500 generations is interesting as it allows us to simulate the ancestral introduction of these genes, and to examine the emergence of subsequent mutations following their introduction. We hypothesise that one of five situations could be observed in response to *def-fmt* reintroduction:

1. We observe no change in genes previously mutated following deletion of *def-fmt*
2. Direct reversions of mutations that appeared following deletion of *def-fmt*
3. Mutations in genes that had not previously been mutated
4. Mutations in genes that had previously been mutated, but which are not reversions
5. No changes occur within a gene at all

As previously discussed (see Chapter 3), under a model of *def-fmt* acting as a toxin-antitoxin gene pair, and being horizontally transferred, we would expect to observe compensatory mutations precluding any ordered loss of the gene pair. In our particular system, we would expect to see compensatory mutations within the translational machinery, and no ordered loss of the *def-fmt* gene pair.

We have now performed whole genome sequencing on 11 lines that have evolved for 3,000 generations with the *def-fmt* pair reintroduced. These lines showed no significant change in fitness in response to *def-fmt* reintroduction (see Chapter 3), making it interesting to investigate how, or indeed whether, these lines have adapted at the genetic level to the presence of these genes. We report a number of parallel mutations across the lines, with some of these seen previously in Lenski's long-term culturing experiments. Importantly, we have seen all five of the aforementioned situations occurring. Some of these mutations have been seen within the same gene in Lenski's long-term culturing experiments and could be attributed to media adaptation. However, some of these mutations have occurred within the translational machinery, and might be important for adaptation to the reintroduction of *def-fmt.*

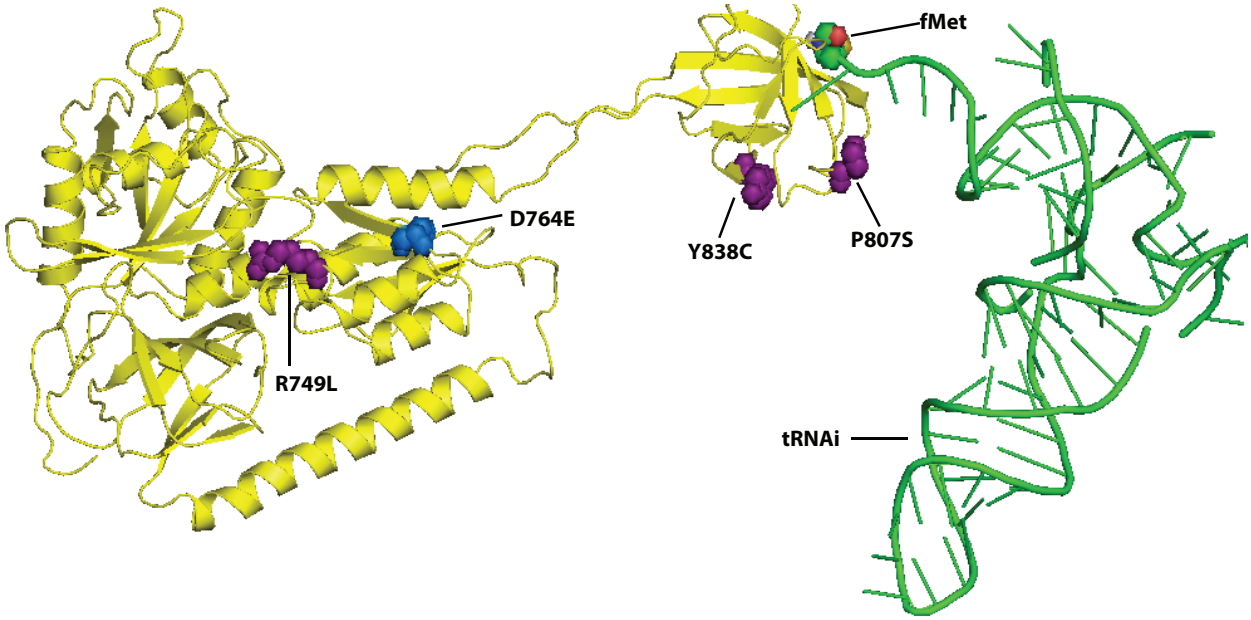


Figure 1. **Parallel mutations in *infB* resulted in changes to initiation factor 2 (IF2) in DF200**. We previously produced an *E. coli* line devoid of the *def-fmt* gene pair and evolved these lines for 1,500 generations (DF200) (Catchpole 2014). Whole genome sequencing of these lines revealed three separate mutations that arose in parallel in 10-11 replicate lines (purple) in *infB*, which codes for IF2 (yellow). One of these mutations was observed in the C1 domain of the protein (R749L), in close proximity to a previous mutation observed in Lenski's experiments (D764E, in blue). These mutations might be important in adaptation to culturing conditions. Two of our mutations (Y838C and P807S) were seen in the C2 domain, the domain that fMet-tRNAi (green) is known to bind to IF2 (Sprink *et al.* 2016). Due to their proximity to fMet-tRNAi, these mutations may be important in compensating translation proceeding with unformylated methionine. Image produced in PyMOL (Schrödinger, LLC 2015) by mapping our mutations to an IF2 structure with fMet-tRNAi bound (3JCJ) (Sprink *et al.* 2016).

Methods

Strains and media

All experiments were performed in a DF200 background, a line derived from a REL606 strain (Genotype: F^-, tsx-467(Am), araA230, lon^-, rpsL227(strR), $hsdR^-$, [mal^+](LamS)). This line has evolved in the absence of the formylase (*fmt)* and deformylase (*def)* genes for 1,500 generations, and has accumulated 58 total mutations. A complete list of strains used is outlined in Chapter 2. The REL606 strain was obtained from T. Cooper (University of Houston, Texas), and all strains were grown at 37°C. A knock-in of the *def-fmt* gene pair was then performed in one of the the DF200 lines (line 6), and 11 independent lines were evolved for a further 3,000 generations (see Chapter 3).

Media used for these experiments was Davis Minimal (Difco), supplemented with 2000mg/L dextrose (DM2000) and 2mg/L thiamine. To make solid media, bacteriological agar was added to a concentration of 1.5% w/v (Oxoid). Streptomycin was also used as specified at 100 μg/ml.

Whole genome extraction

Lines of DF200::*def-fmt* day 100 and DF200::*def* day 100 were streaked to single colonies on DM2000 agar with streptomycin, and were grown at 37ºC for 16

hours. A single colony was subsequently used to inoculate DM2000 media supplemented with streptomycin and was grown overnight at 37°C. DNA was extracted from these cultures using the Wizard® Genomic DNA Purification Kit (Promega), following the manufacturers specifications. DNA was quantified using the Nanodrop 1000 Spectrophotometer and Qubit 2.0 Fluorometer.

Sequencing and Analysis

Whole genome sequencing was performed by Macrogen using an Ilumina Miseq and 2x250bp paired end reads. Raw reads were processed in Geneious R9 (Kearse *et al.* 2012), removing adaptors using the BBDuk plug-in (Bushnell 2014). These trimmed reads were then mapped to the REL606 genome using the Bowtie2 plug-in for Geneious (Langmead & Salzberg 2012), and SNP-calling was performed.

IF2 structural modelling

The mutated *infB* sequence was used to create a homology model with SWISS-MODEL (Biasini *et al.* 2014) and the structure of *E. coli* IF2 bound to the ribosome and fMet-tRNAi (3JCJ) (Sprink *et al.* 2016a). This model contained the substitutions P807S, Y838C and S844F, and was visualised using PyMOL v1.7.4.4 (Schrödinger, LLC 2015). The IF2 protein sequence of 30 other bacterial species was also compared to the sequence of the mutated IF2. These reviewed

sequences were obtained from the Swiss-Prot database (The UniProt Consortium 2015), and were aligned in Genious using the MUSCLE plug-in (Edgar 2004).

Results

Compensatory mutations observed in response to def-fmt reintroduction

In order to investigate the initial evolution of *def-fmt*, we have reintroduced the *def-fmt* gene pair back onto the *E. coli* chromosome in lines that have evolved in its absence for 1,500 generations (DF200::*def-fmt*). We subsequently performed a further 3,000 generations of serial culturing with the genes reintroduced (DF200::*def-fmt* day 100), and have performed whole genome sequencing on these 11 independently evolved lines, as well as 11 control lines with a single *def* knock-in (DF200::*def*). The 22 samples were sequenced using a single Illumina Miseq lane, yielding 7.4GB of data over 13.6 million reads. An average of 617,283±82,692 (X±SD) reads were obtained per sample, with each of these reads an average of 248bp±0.43 (X±SD) long. This sequencing data was of high quality, with 86% of reads having a Q30 value, and gave a sequencing depth of 33±4 per genome.

These raw reads were subsequently processed in Geneious, by trimming and cleaning the sequences with the BBDuk plug-in. These trimmed reads were mapped to the REL606 genome using the Bowtie2 plug-in, giving a sequencing depth of 34±5 ($\bar{x}$±SD). This value excludes ten of the control lines which did not map due to contamination. SNP calling was then performed in Geneious to

identify changes in the sequence. This revealed an average of 61±6 ($\bar{x}$±SD) mutations in the eleven experimental *def-fmt* lines, compared to 19 mutations in the *def* control line. These values exclude those mutations that have already been observed previously in the ancestor DF200 line. These lines share 34±0.6 ($\bar{x}$±SD) mutations with the DF200 ancestor, with the *def* control line having 39 mutations in common with the ancestor. Our previous work has revealed 58 mutations in the DF200 line that was ancestor to both DF200::*def-fmt* and DF200::*def*. An outline of our observed mutations is provided in Table 1.

Our data indicates that there have been direct reversion mutations in some of our DF200::*def-fmt* day 100 lines back to the ancestral REL606 SNP, as we previously hypothesised (and highlighted in Table 3). We have also observed results consistent with our other four hypotheses. We have seen no change in regions previously mutated after *def-fmt* deletion, such as the *metZ* promoter. We have also observed some entirely new mutations in genes where mutations have not previously been observed, and a number of these have occurred in parallel (Table 2). Mutations in previously mutated genes, which are not direct reversions has also been seen. In particular, we have seen this occurring in *trmB* and *infB*, and will be discussing these in further detail (see below). Finally, we have seen genes where there have been no mutational changes at all over the course of *def-fmt* loss and reintroduction. Importantly, we have seen no mutational changes in the *def-fmt* region upon reintroduction (Figure 2). This is consistent with our model of *def-fmt* acting as a toxin-antitoxin gene pair, where we would expect to

see compensatory mutations occurring before any ordered loss of the gene pair. All of these mutations are covered in more detail in future sections.

Table 1. Summary of mutations observed in DF200::*def-fmt* day 100 and DF200::*def* day 100 lines

Line	Number of mutations compared to DF200 (new mutations)	Number of mutations shared with DF200 (not present in REL606 ancestor)	Total mutations compared to REL606 ancestor
DF200::*def-fmt* day 100 line 1	58	34	92
DF200::*def-fmt* day 100 line 2	59	35	94
DF200::*def-fmt* day 100 line 3	55	34	89
DF200::*def-fmt* day 100 line 4	64	34	98
DF200::*def-fmt* day 100 line 5	68	34	102
DF200::*def-fmt* day 100 line 6	66	34	100
DF200::*def-fmt* day 100 line 7	61	34	95
DF200::*def-fmt* day 100 line 8	52	34	86
DF200::*def-fmt* day 100 line 9	65	35	100
DF200::*def-fmt* day 100 line 10	60	34	94
DF200::*def-fmt* day 100 line 11	63	34	97
DF200::*def* control	16	39	55

Contamination in control lines

In the control *def* knock-in line, ten of the eleven sequenced lines did not map back to the REL606 genome. To further investigate this, we performed *de novo* assembly of the genomes using Geneious, and used BLAST to identify the observed sequences. These sequences presented as *Salmonella enterica* serovar Typhimurium, with 100% identity for all 11 lines (E value 0.0). As these lines had tested negative for contaminants at day 100, it is likely that contamination occurred during the DNA extraction process. This makes it difficult to establish whether parallel mutations observed in the experimental lines were unique to that line, as there was only one control line to compare the sequencing data to. It will be important to resequence the control lines in order to establish whether mutations observed in the experimental lines are in response to culturing conditions, or the reintroduction of *def-fmt*.

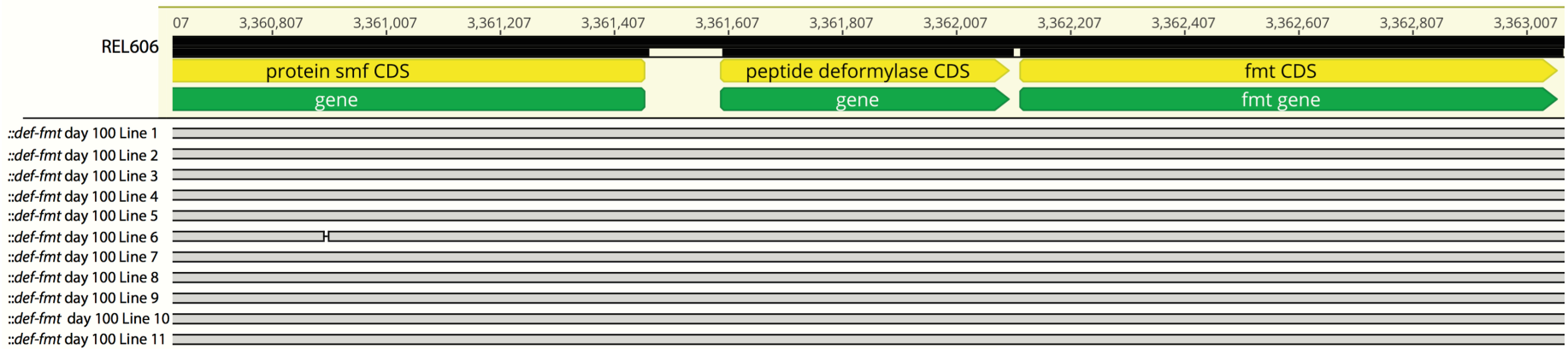


Figure 2. **No changes in the *def-fmt* region were observed after 3,000 generations of long-term culturing.** In order to investigate the effects of *def-fmt* reintroduction, 11 independent lines that have evolved in the absence of *def-fmt* for 200 days were serially cultured for 100 days with the *def-fmt* pair reintroduced onto the chromosome (DF200::*def-fmt* day 100). Whole genome sequencing revealed no changes or loss of the *def-fmt* region in all 11 lines. SNPs or other mutations appear as a gap in the continuous sequence, as can be seen in line 6 in the *smf* gene. This image was created in Geneious, by aligning the sequences of all 11 evolved lines to the REL606 ancestor.

Parallel mutations have occurred across independent lineages

We previously observed parallel mutations in response to the loss of *def-fmt*. While some of these could be attributed to media adaptation, others were in genes encoding the translational apparatus, and, in multiple lineages. These changes were hypothesised to compensate for the loss of *def-fmt*. Now that we have performed a further evolution experiment with *def-fmt* reintroduced (DF200::*def-fmt* day 100), we first examined whether any parallel mutations emerged in our independent lines. We observed 43 parallel mutations across our 11 independently evolved lineages (Figure 3), though only 22 of these mutations were non-synonymous (Table 3). This includes some mutations which have been previously observed in Lenski's long-term culturing experiments (Barrick et al., 2009), and are therefore best explained as adaptations to culturing in minimal media.

We observed parallel mutations in genes that previously picked up mutational changes in response to *def-fmt* loss. These include the genes *infB* and *trmB*, which code for the translational initiation factor 2 (IF2) (Laursen *et al.* 2003) and tRNA (Guanine-N7)-methyltransferase (Liu *et al.* 2008), respectively. The mutational change in *trmB* is a C insertion, but has occurred downstream of a stop codon introduced by a mutation previously observed in DF200 (Figure 4)(Catchpole 2014). The change in *infB* appears to have impacted the way in which fMet-

tRNAi binds to IF2, and will be discussed further. Some of the observed parallel mutations also appear to be reversions back to the REL606 ancestral state.

Table 2. Parallel mutations observed over 6 or more lines in DF200::*def-fmt* day 100 lines. A knock-in of the *def-fmt* gene pair was performed in DF200. These knock-ins were then serially cultured for 3,000 generations, and whole genome sequencing was performed.

Gene	Location	Nuclotide Change	Amino acid change	Function	Lines
secA	113,575	A → G		Preprotein translocase subunit, ATPase (Blanco *et al.* 1996)	11
lpxB	206,444	T → C		Lipid A disaccharide synthase (Crowell *et al.* 1987)	11
	242,196	G insertion		Unknown	8
cstA	631,090	C → T	G178D	Carbon starvation (Schultz *et al.* 1991)	11
speF	702,166	A → G		Ornithine decarboxylase (involved with polyamine biosynthesis) (Chattopadhyay *et al.* 2009)	11
ybgL	725,966	A → G	N6S	Predicted lactam utilisation protein (Finn *et al.* 2014)	11
dinG	827,596	C → T		Helicase (Voloshin & Camerini-Otero 2007)	11
yliE	867,324	T → C		Putative cyclic-di-GMP phosphodiesterase (Rudd 2000)	11
ndh	1,181,148	G → A	A163T	NADH dehydrogenase (Spiro *et al.* 1989)	11
ycjR	1,374,822	A → G	T141A	Unknown (putative sugar transport and catabolism)(Rudd 2000)	10 and control
ynaK	1,420,412	C → G		Unknown, within Rac-like prophage sequence (Rudd 2000)	11
stfR	1,604,898	T → G	K329Q	Side-tail fiber, within Rac-like prophage sequence (Rudd 2000)	11
stfR	1,604,908	A → C		Side-tail fiber, within Rac-like prophage sequence (Rudd 2000)	11
	1,648,450	T → C		Unknown	11
ydgF	1,650,584	A → G	L38P	Drug efflux (Rapp *et al.* 2006)	11
	1,656,675	T → C	-	Unknown	7
malI	1,675,778	T → C		Maltose system regulation – repressor (Reidl *et al.* 1989)	11
ydiB	1,751,480	G insertion	IFS*	Function unclear, but thought to act as a quinate/shikimate	11

				dehydrogenase (Lindner *et al.* 2005)	
ydiY	1,783,302	C insertion	IFS	Unknown – predicted outer membrane protein (Rudd 2000)	11
yeaY	1,868,275	A → T	I28N	Outer membrane lipoprotein (Keseler *et al.* 2013)	11
	1,995,669	C insertion		Unknown	7
yfeO	2,445,091	C insertion	FS*	Putative transport protein (The UniProt Consortium 2015)	9
yfeS	2,466,972	T → C		Unknown	11
alaS	2,716,283	G → C		Alanine-tRNA ligase (Putney *et al.* 1981)	11
trmB	2,987,996	C insertion	IFS	tRNA (Guanine-N7)-methyltransferase (The UniProt Consortium 2015)	11
	3,222,311	C → T		Unknown	11
infB	3,248,718	G → A	S844F	Translation initiation factor 2 (Wienk *et al.* 2012)	7
gltB	3,290,633	G → A	R225H	Glutamate synthase (Pahel *et al.* 1978)	8
dcuD	3,303,590	T → C	V275A	Dicarboxylate uptake (Janausch & Unden 1999)	11
pckA	3,461,693	T → C		Phosphoenolpyruvate carboxykinase (Medina *et al.* 1990)	11
rhsB	3,551,397	T → C		Unknown	6
yhjA	3,598,030	C insertion	IFS	Predicted cytochrome C peroxidase (Partridge *et al.* 2007)	11 and control
trbC	3,785,850	T → C	L314P	Conjugal transfer protein (The UniProt Consortium 2015)	11
ibpA	3,827,577	C → T		Heat-shock protein (Kuczyńska-Wiśnik *et al.* 2002)	11
yigG	3,967,480	A → G	I50T	Predicted inner membrane protein (Rudd 2000)	11
hslU	4,101,184	T → C	H16R	ATPase involved in degradation of misfolded proteins (Ramachandran *et al.* 2002)	11
	4,228,268	C → T		Unknown	11
yjfY	4,402,785	C → T	R54H	Unknown	11
pyrB	4,455,024	C → T	E217K	Aspartate carbamoyltransferase (Pauza *et al.* 1982)	7
pyrL	4,455,696	A deletion	FS	PyrBI operon attenuator (Keseler *et al.* 2013)	11
yjiL	4,544,826	A insertion	IFS	Unknown	11
yjiV	4,553,398	A → G		Unknown	11
slt	4,620,776	A → G		Soluble lytic transglycosylase (Betzner & Keck 1989)	11

*FS: frameshift mutation, IFS: inactivating frameshift

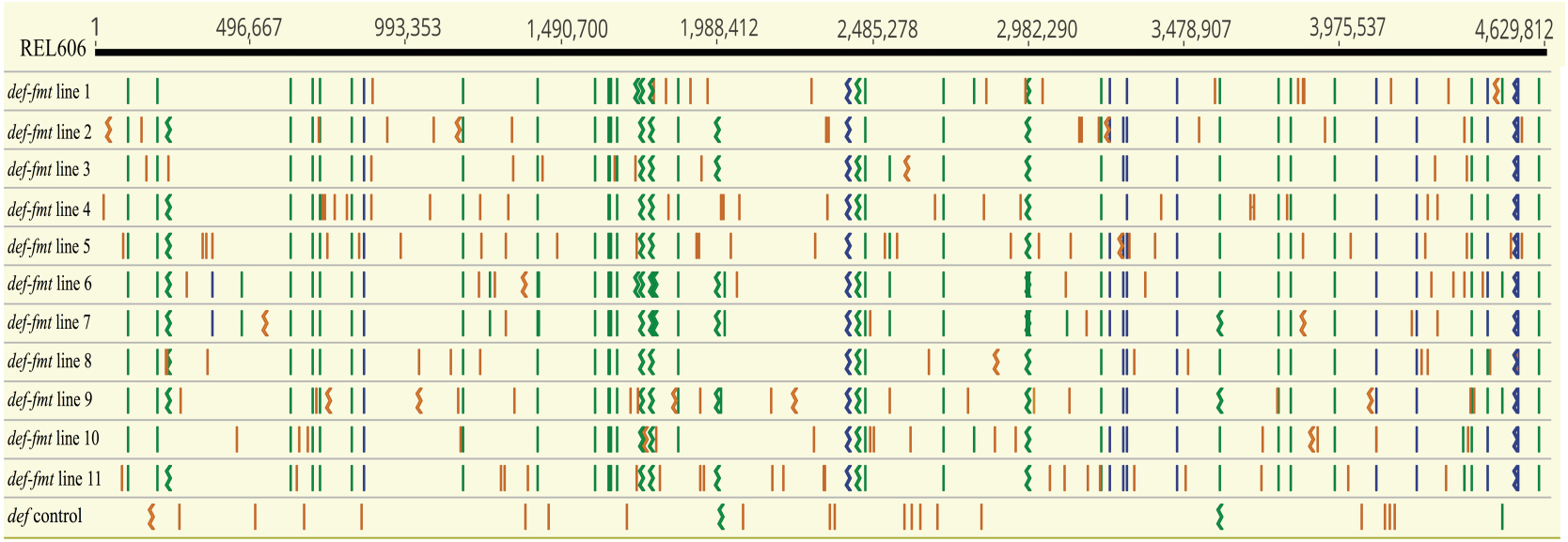


Figure 3**. Parallel mutations have been observed across multiple DF200::*def-fmt* day 100 lines.** 11 lines of DF200::*def-fmt* were cultured independently for 100 days, and whole genome sequencing was performed. These reads were then mapped to the REL606 genome (NC_012967) using the Bowtie2 plugin in Geneious. Parallel mutations observed over multiple lines are indicated by green lines. Some of these have been observed in Lenski's long-term culturing experiments (blue), while some mutations appear to be independent in each of the lines (orange). SNPs are indicated by a single dash, while indels are indicated by a zig-zag line. Those mutations which have already been observed in the DF200 lines have been removed from this image. This image was created in Geneious.

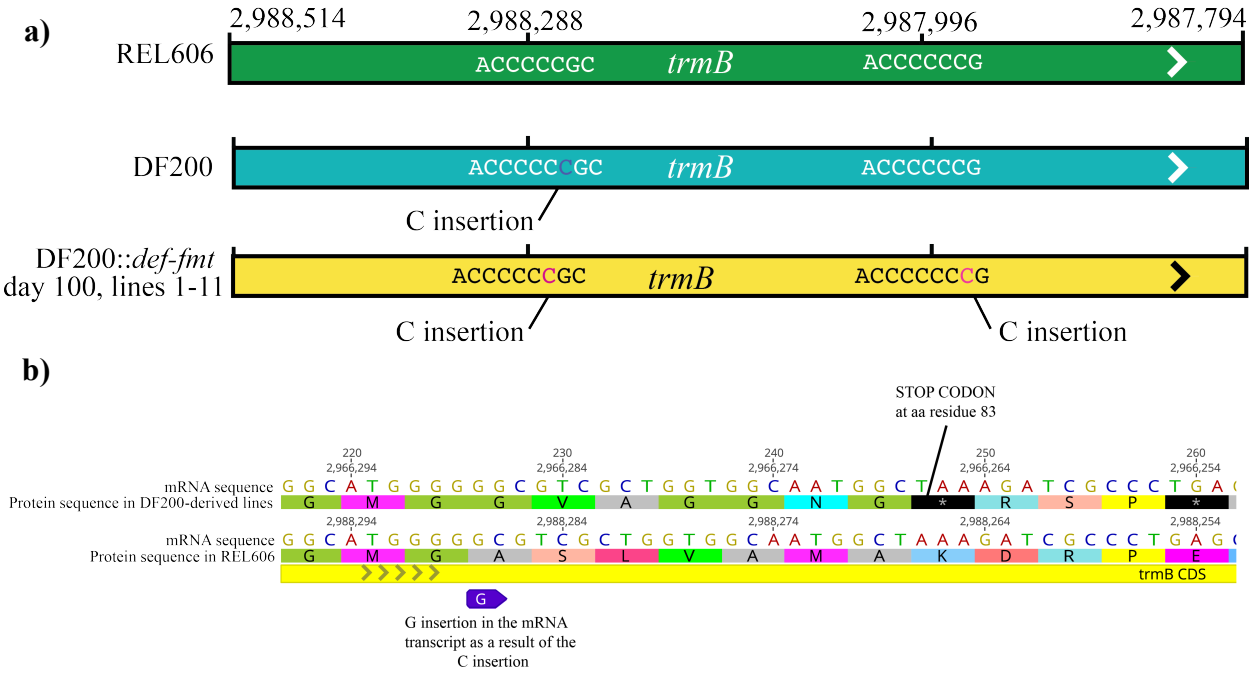


Figure 4. **Mutations in *trmB* appear to result in protein truncation**. **a)** 11 lines of DF200::*def-fmt* were cultured independently for 100 days, and whole genome sequencing was performed, revealing parallel mutations in *trmB*. A mutation in this region had been observed following adaptation to *def-fmt* loss (DF200) (Catchpole 2014). This mutation in DF200 resulted in the insertion of a C at position 2,988,288 in the genome, which resulted in a premature stop codon in the protein sequence, 8 residues after the mutation. We have now observed a second mutation that results in the insertion of a C in DF200::*def-fmt* day 100 lines at 2,987,996. **b)** The mutation in DF200 (pictured as a G on the mRNA transcript) resulted in the insertion of a premature stop codon (indicated by an asterisk in the protein sequence). The new mutation has occurred downstream of the introduced stop codon, and is not likely to have an effect, as the protein is expected to have been truncated as a result of the previous mutation. This further mutation is consistent with pseudogenisation of a previously inactivated gene.

Reversion mutations observed in response to def-fmt reintroduction

As mentioned above, some of the parallel mutational changes appear to be reversion mutations back to the ancestral REL606 state (Figure 5, Table 3). These are situations where a mutation has been observed in the DF200 lines after *def-fmt* loss, and after further evolution experiments post-*def-fmt* reintroduction, we have observed a direct reversion back to the REL606 wildtype ancestral state. Under our experimental conditions, we might expect some reversion mutations, as we have reintroduced *def-fmt* into DF200 after 1,500 generations without the gene pair. A number of our reversion mutations are shared with the single knock-in control (DF200::*def*). Their presence in the control line would indicate that these reversion mutations are not important for adapting to the presence of formylated methionine. Mutations that are present in our control line cannot be entirely disregarded, however, as these lines are expressing peptide deformylase, which is part of the formylation process.

A small number of these reversion mutations are only seen in the experimental lines, including mutations in *yaiT, ybaL, xdhA and, yecO* which were observed in all 11 experimental lines, and a reversion mutation in a hypothetical protein (ECB_00735) that occurred in ten of the lines. It is difficult to conclude what the role of some of these mutations might be, as they occur in poorly characterised genes. To further investigate the function of these poorly characterised genes (*yaiT* and ECB_00735) we have searched the protein sequences in Pfam (Finn *et*

al. 2014) and Phyre2 (Kelley *et al.* 2015). Searches on Pfam for *yaiT* provided hits for an autotransporter (e-value 5.2e-45) and a far weaker hit for pertactin (a type of outer membrane protein, e-value 2.5e-05). Upon searching the gene in Phyre2, it was found that in other *E. coli* strains, this gene has been annotated as an outer membrane autotransporter. Protein structure prediction in Phyre2 provided a pre-cleavage structure of an autotransporter (100% confidence and 30% coverage). As a result of this, it can be concluded that *yaiT* is likely involved in protein transport. For ECB_00735, however, no significant hits were found on Pfam, and Phyre2 was not able to create a structure with a high confidence (superfamily: bacterial enterotoxin, 37% confidence and only 14% coverage). It is therefore difficult to conclude what role a reversion in ECB_00735 might be playing. To establish whether this protein is important for adapting to the absence of formylation, we could perform a knock-in of our mutated sequence into a *def-fmt* knock-out line. As these lines normally grow extremely slow, a benefit to the cell (in this case, an increase in fitness) in the absence of formylation should be possible to see.

A number of these mutations did not occur in parallel in the original experiment adapting to *def-fmt* loss (Catchpole 2014), instead only occurring in the one ancestral line that was used for the *def-fmt* knock-ins. While a reversion mutation in *yecO* appears to be involved with translation, this mutation was not observed in parallel previously. This might make it difficult to conclude whether the parallel reversion back to the REL606 ancestral state was important in adaptation

to *def-fmt* reintroduction. *xdhA,* while also lost in parallel in all experimental lines, only occurred in one line in the initial evolution experiment adapting to *def-fmt* loss. However, if there are multiple routes by which *E. coli* can adapt to *def-fmt* loss, mutations in *yecO, yaiT, ybaL* and ECB_00735 might be important in adapting to *def-fmt* reintroduction, even if such mutations were not in parallel in the DF200 lines after *def-fmt* loss.

Table 3. Reversion mutations observed in *def-fmt* lines. These are mutations that were observed in DF200, but were not seen after *def-fmt* reintroduction and a further 3,000 generations of long-term culturing.

Gene	**Location**	**DF200 mutation**	**Nucleotide in reverted lines**	**Function**	**Lines**
ecpD	158,672	G → A	G	Pillin porin protein (Raina *et al.* 1993)	11 and control
	292,850	A → G	A	Upstream *ykgD*	11 and control
yaiT	360,605	G insertion	G deletion	Unknown	11
ybaL	474,482	A → G	A	Inner membrane protein (The UniProt Consortium 2015)	11
ECB_00735	794,714	G insertion	G deletion	Unknown – within prophage 434 sequence	10
ECB_00840	897,556	A → T	A	Unknown	11 and control
ECB_00843	898,997	T → C	T	Phage baseplate assembly protein (within 186-like prophage sequence)	11 and control
ndh	1,181,315	A → G	A	NADH dehydrogenase (Spiro *et al.* 1989)	11 and control
yecO	1,931,520	G → A	G	tRNA (cmo5U34)-methyltransferase (The UniProt Consortium 2015)	11
cheB	1,946,971	C insertion	C deletion	Chemotaxis MCP protein-glutamate methylesterase (Rudd 2000)	11 and control

yehV	2,166,521	T → C	T	Putative transcriptional regulator (The UniProt Consortium 2015)	11 and control
yeiG	2,195,249	C deletion	C insertion	S-formylglutathione hydrolase (Gonzalez *et al.* 2006)	10 and control
dsdX	2,412,185	T → C	T	D-serine transporter (Anfora & Welch 2006)	11 and control
	2,663,274	A → T	A	Upstream *yfiR* (unknown function)	11 and control
xdhA	2,887,404	C → T	C	Xanthine dehydrogenase (Rudd 2000)	11
	2,891,925	T → C	T	Between *ygeV* (putative transcriptional regulator) and *ygeW* (predicted carbamoyltransferase) (Keseler *et al.* 2013)	11 and control
yeeS	3,007,723	A → G	A	Predicted DNA repair protein (Keseler *et al.* 2013)	11 and control
glmM	3,259,228	A → G	A	Phosphoglucosamine mutase (Mengin-Lecreulx & van Heijenoort 1996)	11 and control
nikE	3,549,138	G → C	G	Periplasmic binding-protein-dependent nickel transport system (Navarro *et al.* 1993)	11 and control
dppF	3,632,502	C → G	C	Dipeptide transport ATP-binding protein (The UniProt Consortium 2015)	11 and control
yjbD	4,210,573	G insertion	G deletion	Unknown	11 and control
gpmB	4,622,495	C → G	C	Probable phosphoglycerate mutase (The UniProt Consortium 2015)	11 and control

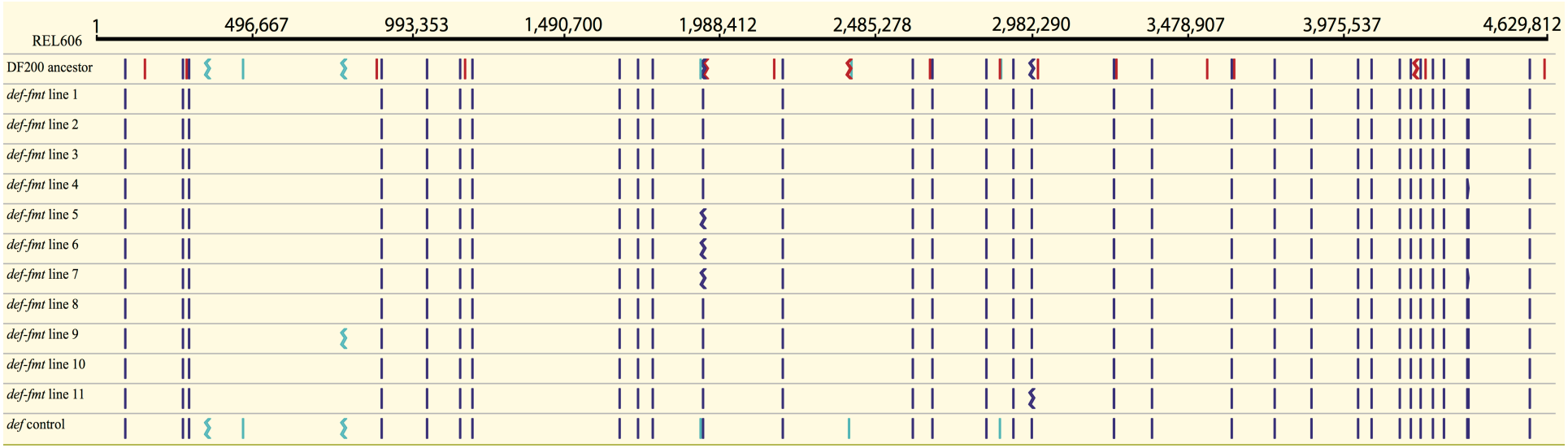


Figure 5. **A map of reversion mutations in DF200::*deffmt* day 100.** A knock-in of the *def-fmt* gene pair was performed in DF200, and were serially cultured for 3,000 generations (100 days). Whole genome sequencing was subsequently performed. A number of mutations that were observed in DF200 lines, but not in the REL606 control line have been retained (purple), while some mutations appear to have reverted back to the REL606 ancestral state. While some of these appear to have also reverted in the *def* control line (light blue), a number of these appear to have only reverted in the experimental lines (red).

Mutations in infB in region of fMet-tRNAi binding

Previously we observed three mutations in *infB* when adapting *def-fmt* knock-out lines to the absence of the gene pair: R749L, P807S and Y838C (Catchpole 2014). Two of these mutations, P807S and Y838C, were found in the C2 domain of initiation factor 2 (IF2) (Figure 6), an area fundamental to the binding of the fMet-tRNAi (Wienk *et al.* 2012; Sprink *et al.* 2016). The third mutation, R749L, was found in another region of IF2, and mutations in this region have been previously linked to Lenski's long-term culturing experiments (Barrick *et al.* 2009). This means that this third mutation is most likely associated with adaptation to culturing conditions, so seems unlikely to have occurred specifically in response to the loss of *def-fmt*. There is one difference between our specific culturing conditions and those used by Lenski. We have used eight times as much glucose in our experiments (DM25 vs DM200) (Lenski *et al.* 1991), meaning that, while glucose is limited, it is not as limited as in Lenski's experiments. All other conditions were similar. 11 parallel lines were grown at 37°C in minimal media, with serial culturing performed once every 24 hours, meaning that mutations in our experiments that were also observed in Lenski's experiments are most likely a result of media adaptation.

Whole genome sequencing of our DF200::*def-fmt* day 100 lines has now revealed a further mutation in *infB*, in 7 of the 11 lines, resulting in a S844F change. In order to investigate the impact of this mutation, we have created a homology

model of IF2 using the structure for *Escherichia coli* ribosome-bound IF2 (3JCJ). The observed mutations in *infB* were then mapped to this model, to paint the picture of our mutations within the IF2 structure. The S844F mutation maps to the C2 domain of the IF2 protein, the domain where previous mutations have been observed in response to the loss of *def-fmt*. This mutation has occurred very close to where tRNAi-fMet appears to bind (Figure 7), with a change from a hydrophilic to a hydrophobic amino acid residue. We next compared the sequence of the C2 domain of IF2 across 8 bacterial phyla, where reviewed IF2 sequences were available, in order to establish whether the changes we are seeing are seen in other species, or whether this is a change that might be unique to our lines (Figure 8). We also included one species of mollicute in this list (*Mycoplasma hyponeumoniae*), as this species has lost the formylase-deformylase gene pair (Grosjean et al., 2014). This revealed that, for P807S, the serine residue is observed in other bacterial species, while for the other two changes we have observed, Y838C and S844F, we do not see the cysteine and phenylalanine residues present in the wide range of bacterial species we have compared our line to. This suggests that these particular changes are uncommon, and might be unique to our experimental conditions.

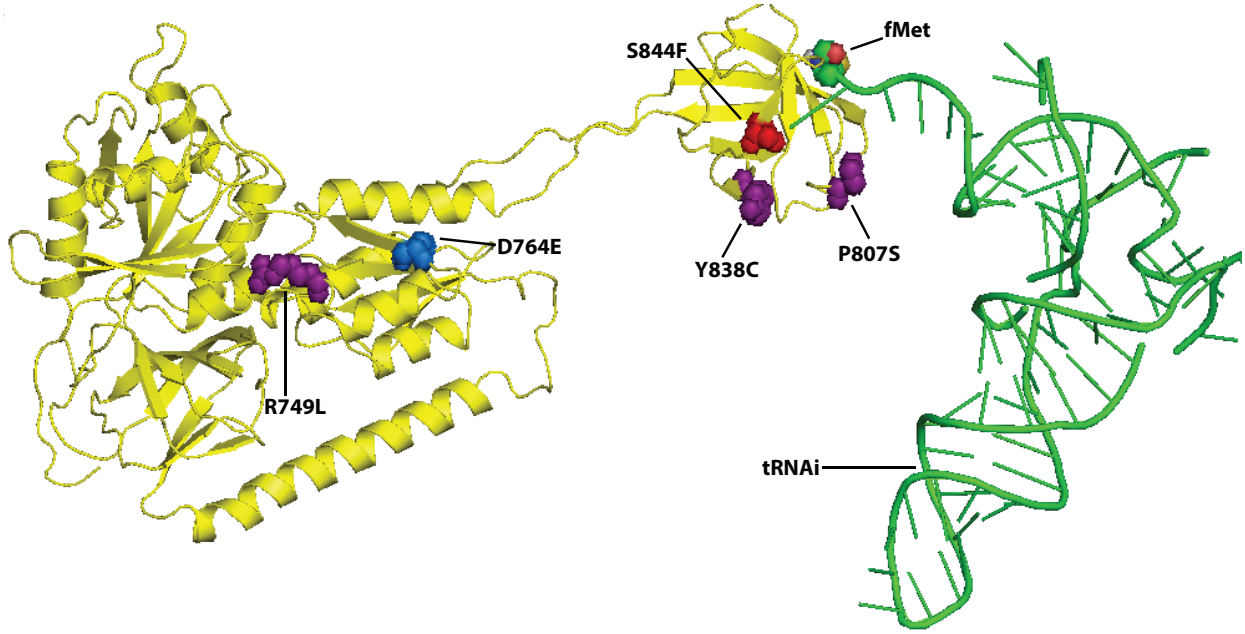
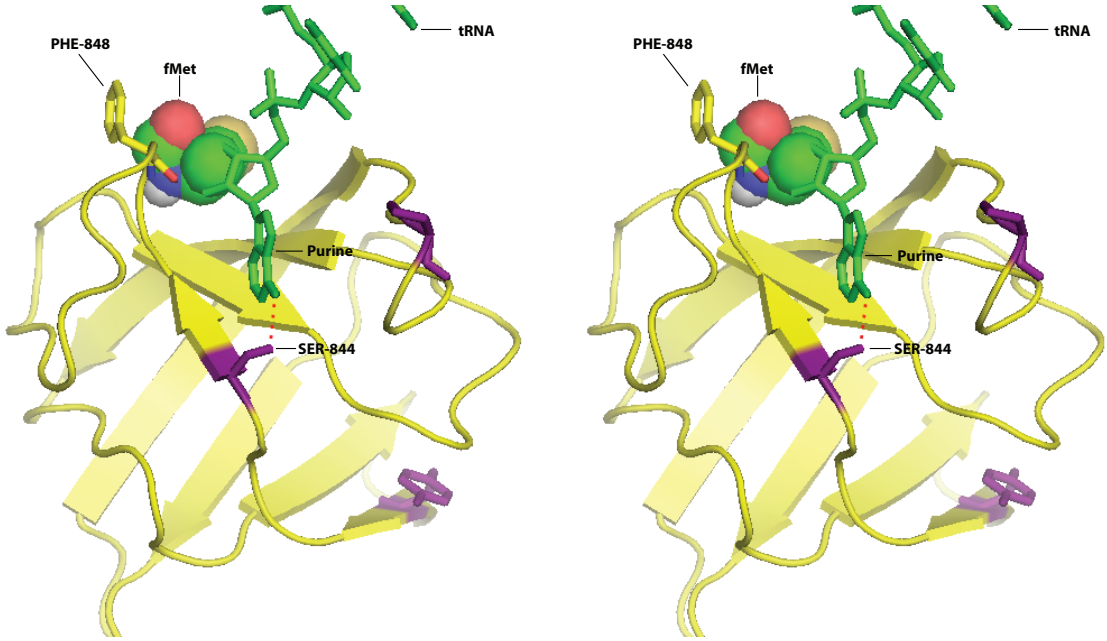
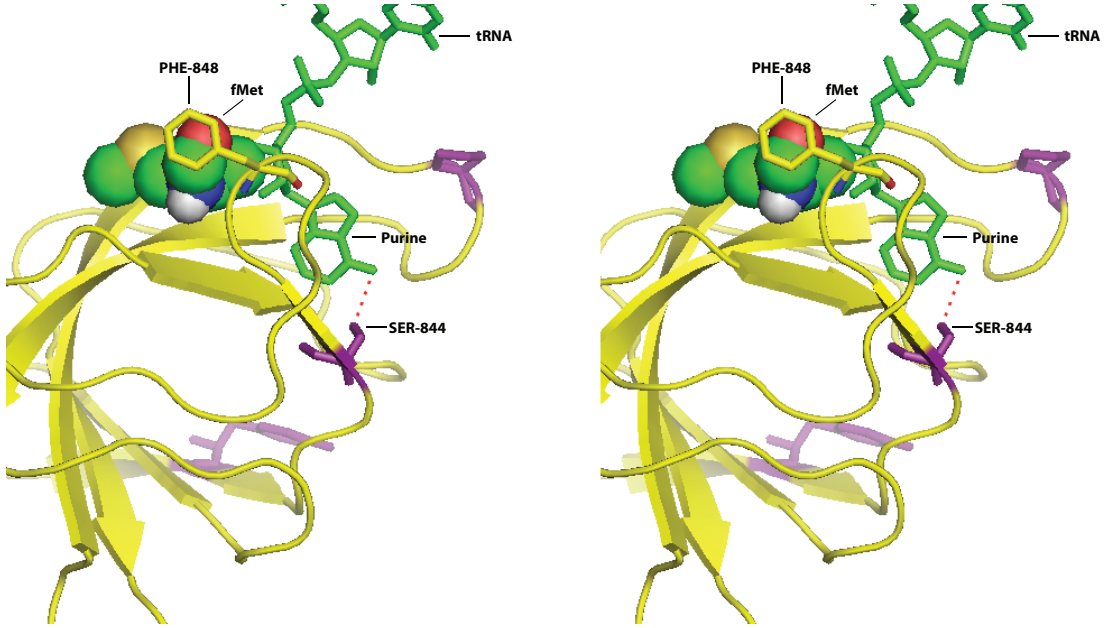


Figure 6. **A number of mutations have been observed in *infB* in both the DF200 and DF200::*def-fmt* day 100 lines, leading to changes of amino acid residues in initiation factor 2 (IF2).** As previously mentioned in Figure 1, one of these mutations in DF200, R749L, is in a region where previous mutations have been observed in Lenski's experiments (D764E, in blue). Two other mutations in DF200, Y838C and P807S (purple), are in the C2 domain, a domain important for tRNAi-fMet (green) binding. In 7 of our 11 DF200::*def-fmt* day 100 lines, we have seen one further mutation within the C2 domain, S844F (red). These three mutations in the C2 domain are in close vicinity of fMet-tRNAi. This image was generated in PyMOL *(Schrödinger, LLC 2015)* mapping our mutations to an IF2 structure with fMet-tRNAi bound inside the 70S ribosome (3JCJ) *(Sprink* et al. *2016a)*.

a)


PHE-848
fMet
tRNA
Purine
SER-844
PHE-848
fMet
tRNA
Purine
SER-844

b)


PHE-848
fMet
tRNA
Purine
SER-844
PHE-848
fMet
tRNA
Purine
SER-844

c)

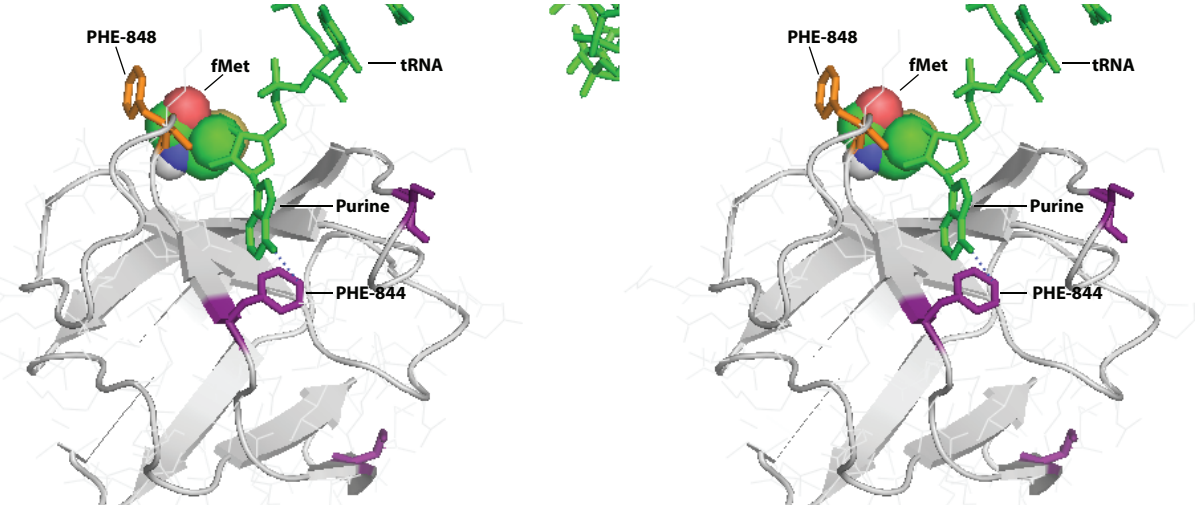

PHE-848
fMet
tRNA
Purine
PHE-844
PHE-848
fMet
tRNA
Purine
PHE-844

Figure 7. **Mutations in *infB* result in changes in IF2 (purple) within close proximity of fMet-tRNAi.** In our DF200::*def-fmt* day 100 line, we have seen a S844F change in 7 of our 11 lines. **a)** The SER-844 residue appears in close vicinity of a purine on the fMet-tRNAi. **b)** This mutation is only five residues away from the PHE-848 residue which is in direct proximity of the formyl group of fMet-tRNAi. **c)** In order to visualise our mutation, we have created a homology model of the mutated IF2 sequence using SWISS-MODEL *(Biasini* et al. *2014)*, and the structure of *E. coli* IF2 bound to the ribosome and fMet-tRNAi (3JCJ) *(Sprink* et al. *2016a)*. Our newly observed mutation is a change from a hydrophilic to a hydrophobic residue, and due to its close proximity to fMet-tRNAi, might be affecting the way in which IF2 binds to the initiating methionine. Images were generated in PyMOL *(Schrödinger, LLC 2015)*.

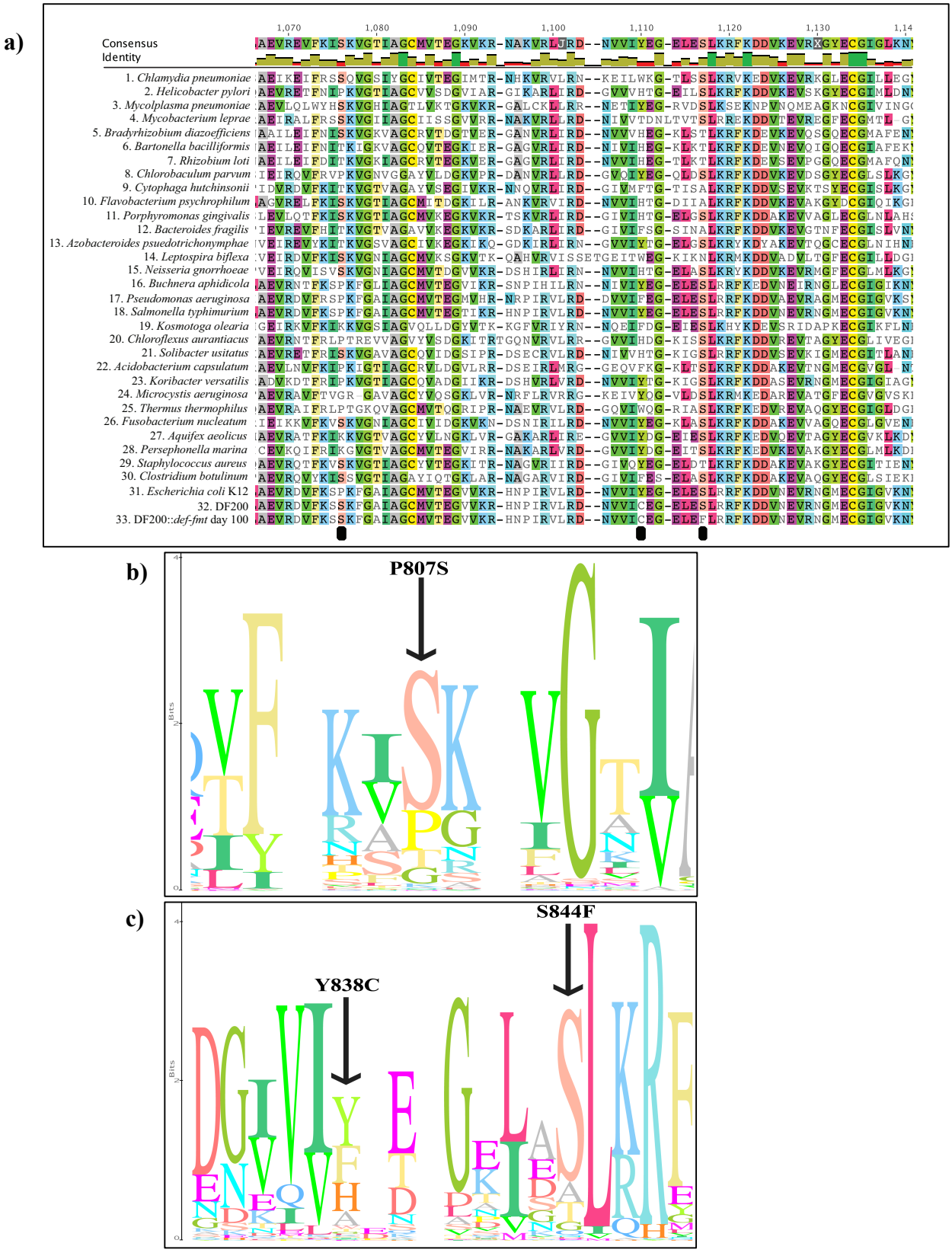

a)
1,070
1,080
1,090
1,100
1,110
1,120
1,130
Consensus
Identity
1. Chlamydia pneumoniae
2. Helicobacter pylori
3. Mycolplasma pneumoniae
4. Mycobacterium leprae
5. Bradyrhizobium diazoefficiens
6. Bartonella bacilliformis
7. Rhizobium loti
8. Chlorobaculum parvum
9. Cytophaga hutchinsonii
10. Flavobacterium psychrophilum
11. Porphyromonas gingivalis
12. Bacteroides fragilis
13. Azobacteroides psuedotrichonymphae
14. Leptospira biflexa
15. Neisseria gnorrhoeae
16. Buchnera aphidicola
17. Pseudomonas aeruginosa
18. Salmonella typhimurium
19. Kosmotoga olearia
20. Chloroflexus aurantiacus
21. Solibacter usitatus
22. Acidobacterium capsulatum
23. Koribacter versatilis
24. Microcystis aeruginosa
25. Thermus thermophilus
26. Fusobacterium nucleatum
27. Aquifex aeolicus
28. Persephonella marina
29. Staphylococcus aureus
30. Clostridium botulinum
31. Escherichia coli K12
32. DF200
33. DF200::def-fmt day 100
b)
P807S
Bits
c)
S844F
Y838C
Bits

Figure 8. **Protein alignment reveals mutations to residues not seen in other phyla**. **a)** The IF2 protein sequences of 30 different bacterial species were obtained from the Swiss-Prot database, selected from the eight bacterial phyla with reviewed IF2 sequences. These sequences were compared to the *Escherichia coli* K12, DF200 and DF200::*def-fmt* day 100 mutated sequences. These were then aligned using the MUSCLE plug-in in Geneious. The mutations we have observed are indicated by a black dash. While one of the changes seen in IF2, P807S, appears to be present in other bacterial lineages, the other two mutations we have seen, Y838C and S844F, were not observed across other lineages. This would indicate that these two changes are not common for the initiation factor, and might be important to our experimental conditions, where we have evolved *E. coli* without *def-fmt* and have subsequently reintroduced the gene pair and evolved the lines for a further 3,000 generations (100 days). **b)** To establish whether the residues we're observing mutations in are conserved over a larger group of bacteria, we have generated a sequence logo by aligning the 557 bacterial IF2 sequences available in the Pfam database in Geneious. Of our changes, the P807S change was the least conserved, with more bacterial sequences possessing our mutated reside (serine). **c)** For the Y838C, tyrosine (Y) is the most common residue seen, but cysteine (C) is seen in 2 of the 557 sequences. For the S844F change, serine is very well-conserved across this residue, and phenylalanine (F) is not observed in any of the 557 sequences. This would indicate that this residue is unusual at this position of the protein.

Discussion

Our sequencing data has revealed all five possible aforementioned mutational outcomes. We have seen examples of no mutational changes occurring in genes previously observed to be mutated in DF200, such as *metZ*. Direct reversion mutations have also been seen, where there has been a reversion back to the REL606 wildtype ancestral state in our DF200::*def-fmt* day 100 lines. A number of these mutations were observed in parallel in response to *def-fmt* reintroduction. We have also seen situations where mutations have been seen in previously mutated genes, but these are not direct mutations of previous mutations. This was seen in *infB* and *trmB*, both genes associated with translation. Further to this, we have seen a number of new parallel mutations in genes that were not previously seen to have mutations. Some of these could be attributed to culturing conditions as they had been previously observed in Lenski's long-term culturing experiments. Finally, we have seen genes with no mutational changes at all, in particular, the *def-fmt* region has observed no mutational changes, important to our toxin-antitoxin model. The relevance of our observed mutations will now be discussed in detail.

It is important to note that in our experimental context, mutations observed in either our control line or in Lenski's previous long-term culturing experiments are not likely to be important to our experimental conditions. These mutations can be attributed to adaptation to culturing conditions. Our *def* control, in

particular, is interesting as it allows us to solely attribute any mutations in our experimental line to the presence of formylation (as the only difference between the two lines is the absence of the *fmt* gene in the control line).

1. Mutations in genes that had not been previously mutated

We have observed a number of new parallel changes in our DF200::*def-fmt* day 100 lines, with a number of these having been previously seen in the same gene in Lenski's long-term evolution experiments, and are therefore likely to be important in adaptation to media conditions (Barrick *et al.* 2009). This includes genes involved with a variety of different functions, spanning roles from glutamate synthase (*gltB*)(Pahel *et al.* 1978), to decarboxylate uptake (*dcuD)*(Janausch & Unden 1999), degradation of misfolded proteins *(hslU)*(Ramachandran *et al.* 2002) and aspartatecarbamoyl transferase (*pyrB)*(Pauza *et al.* 1982). We have also observed parallel mutations in *yjiL*, which is a gene of unknown function. This makes it difficult to conclude what role it might be playing in adaptation to the environment, especially as searches in Pfam and Phyre2 returned no significant results for this sequence. The fact that certain mutations are occurring in Lenski's long-term culturing experiments, or in our control line which is not formylating, would indicate that these mutations might be important for adaptation to media conditions. Lenski has attributed many of his mutations to adaptation to the limiting carbon source within the minimal media used (Lenski 2011). Our conditions were slightly different to

those imposed by Lenski, in that we used eight times as much glucose, and only performed serial culturing for 2,000-3,000 generations compared to Lenski's tens of thousands of generations. However, the other conditions used were the same, making it likely that these mutations are adaptations to minimal culturing conditions, as opposed to *def*-fmt reintroduction. Further to this, we have seen two mutations which, while not previously seen by Lenski, were observed in our control line: *ycjR* and *yhjA*. The presence of these mutations in our control line would indicate that these mutations are not exclusively important for adaptation to the presence of formylation. However, as we observed contamination in ten of our eleven control lines, and have therefore only received information on one of our controls, it is possible that we might observe some of our mutations that we consider to be exclusive to our DF200::*def-fmt* day 100 lines in our other control lines. To establish whether our mutations are truly important for adaptation to formylation, it will be important to resequence our ten DF200::*def* control lines in the future, to create a true picture of which genes are important in adapting to *def-fmt* reintroduction.

We have also observed a number of parallel mutations in genes, which have not been previously associated with Lenski's experiments, and were not seen in our control lines. One of these mutations possesses functions that could be associated with culturing conditions, *cstA*: a carbon starvation gene. While mutations in *cstA* have not been previously seen by Lenski, the function of this gene would indicate

that it could be important in responding to an environment with a limited carbon source, such as the minimal conditions we are using.

We have also seen fourteen mutations resulting in changes in amino acid residue that were parallel in our lines. Three of these mutations, *ydiB*, *ydiY* and *yhjA* have resulted in frameshifts that introduce premature stop codons into the protein sequence. Our previous results would indicate that there has been no loss of fitness as a result of inactivation of these genes (see Chapter 3), and that these genes are not important in our specific culturing conditions. In particular, *yhjA* codes for a predicted cytochrome c peroxidase, a transmembrane protein that has found to only be expressed in conditions of oxygen starvation (Partridge *et al.* 2007). We have not subjected our lines to conditions of oxygen starvation, meaning that deactivating frameshift mutations would be tolerated under our specific culturing conditions. The function of many of these genes where we have seen parallel mutations do not appear to be directly associated with fMet-tRNAi^{fMet}, and may therefore not be relevant to *def-fmt* reintroduction. As many proteins associated with translation are still being discovered, we are not able to definitively conclude that these mutations are not occurring in response to *def-fmt* reintroduction. To establish whether these mutations are relevant in the future, we may perform two further experiments: resequencing of control lines and knock-in experiments. As previously discussed, some of these parallel changes may have also occurred in our other ten control lines, but due to contamination we have been unable to see this. Resequencing these lines will aid in establishing

how important these parallel mutations are to adapting to *def-fmt* reintroduction. Performing knock-ins of these genes into our DF200::*def-fmt* day 0 line may also elucidate whether these mutations are having an impact on our lines. One of our parallel mutations (in *yjfY*) is in a gene of unknown function, with searches in Pfam giving no significant results and searches in Phyre2 returning results for a YdgH-like protein (100% confidence, 97% coverage). *ydgH* is also a gene of unknown function, but in *Salmonella,* the first 22 amino acid residues resemble a signal peptide, and it is thought this signal indicates periplasmic localisation (Eletsky *et al.* 2014). The mutation we have observed in parallel may therefore play a role in bacterial cell membranes, and the relevance of such mutations in the context of our experiment will be further discussed in section 3 of the discussion.

2. No change observed in genes previously mutated following deletion of *def-fmt*

In this thesis, we have seen situations where previous mutations observed in DF200 were unchanged. This included the *metZ* promoter region, where a previous mutation was thought to increase $tRNAi^{Met}$ expression. This was tested using a GFP reporter system, where increased expression of GFP was seen during stationery phase under the mutated *metZ* promoter. Based on these results, and the hypothesised impact of the mutations we observed in IF2, we concluded that the higher expression of *metZ* may help compensate for the decreased affinity of

IF2 for binding to unformylated methionine (Catchpole 2014). That we observed no further mutations in this region may indicate that while selection for increasing the concentration of charged tRNA is strong if the affinity is lowered, re-emergence of high-affinity would not necessarily result in strong selection against the now-elevated tRNA levels. However, as previously discussed in the introduction, these results in a GFP reporter system may have limitations due to the fact that we have not explicitly established whether this mutation has actually results in increased fitness in our lines. This makes it difficult to conclude how large an impact this mutation is having on adaptation to *def-fmt* loss. It will be important to analyse the effects of mutations in the *metZ* promoter in an evolved DF200 line where we knock-in the wildtype sequence. This will allow us to establish whether mutations in the *metZ* promoter are resulting in increased fitness in the absence of formylation.

3. Direct reversions of mutations that appeared following deletion of *def-fmt*

We have also observed a number of reversion mutations in parallel within our lines, with a number of these also occurring in our *def* control line (Table 3). Of these reversion mutations, only four are seen exclusively in the experimental lines. These mutations include changes in both the *yecO* and *xdhA*, coding for tRNA (cmo5U34)-methyltransferase and xanthine dehydrogenase respectively (The UniProt Consortium 2015). *yecO* has S-adenosylmethionine

methyltransferase activity (The UniProt Consortium 2015), like *trmB*, a gene where we have observed inactivating frameshift mutations. tRNA (cmo5U34)-methyltransferase is involved with the conversion of 5-methoxyuridine to uridine-5-oxyacetic acid at the wobble position of tRNA (position 34 of the tRNA) (Nasvall 2004). The presence of uridine-5-oxyacetic acid is considered important for the decoding of codons ending in G, and experiments in *Salmonella enterica* have shown this to be essential with proline, valine and alanine codons ending in G (Nasvall *et al.* 2007). Methionine tRNA is modified by 4-acetylcytidine (ac^4C) in the wobble position, a modification that is unique to the methionine residue, and ensures differentiation from the similar isoleucine residue (Stern & Schulman 1978; Cantara *et al.* 2013). Genome-wide screens of uncharacterized genes in *E. coli*, have concluded that the *ypfl* gene likely codes for this modification, acting as a tRNA(Met) cytidine acetyltransferase (Ikeuchi *et al.* 2008). However, research that specifically implicates this modification in the initiating fMet-$tRNAi^{fMet}$ could not be found, so it is therefore uncertain whether this modification occurs exclusively on elongator Met-tRNA. Searches through MODOMICs, a RNA modification database (Machnicka *et al.* 2013), only provides a single entry for *Escherichia coli* Met-tRNA, and does not have an entry for the initiating fMet-tRNAi. If this modification does also occur on formylated methionine, it is likely that *yecO* is not implicated in modification of fMet-$tRNAi^{fMet}$. This would mean that mutations in this region are likely to not be implicated in adaptation to formylation loss or reintroduction. However, in order to fully establish whether this is the case, further experiments isolating the

effect of this mutation by knocking the genes into a wildtype REL606 line with *def-fmt* knocked out could be used to see if the presence of this mutation is important in adaptation to *def-fmt* loss, and whether the subsequent reversion is due to *def-fmt* reintroduction. However, as *def-fmt* knock-outs are difficult to grow, this experiment might be difficult to complete. A knock-in of the wildtype REL606 sequence into our evolved double knock-outs (DF200) will aid in determining whether this mutation is important for adaptation to *def-fmt* loss. If we observe any loss of fitness as a result of introducing the wildtype sequence, then it is likely that this mutation is important in adapting to *def-fmt* loss.

A further reversion in *yaiT* was seen in parallel previously, but function of this protein is unknown. Further analyses of the *yaiT* sequence in Pfam and Phyre2 have established that this gene likely encodes a protein that is involved in protein transport. We have also seen a reversion mutation in *ybaL*, an inner membrane protein. Mutations in proteins involved in protein transport or membrane proteins, such as *yaiT* and *ybaL,* could be important for adapting to the presence to the loss of *def-fmt* (Catchpole 2014). This is because membrane proteins produced in the ribosome are targeted to the membrane before the polypeptide sequence is complete (Bornemann *et al.* 2008), and as a result of this, these proteins might not have the formyl group removed by peptide deformylase before the protein reaches the membrane. Further to this, the *N*-formylmethionine is known to be retained in some membrane proteins (Milligan & Koshland 1990; Coleman 1992; Guan *et al.* 2011). This means that a reversion mutation in *ybaL* or *YaiT* could be

important for compensating for a formyl group being retained on the protein when it reaches the membrane. In this situation, we would expect that the first mutation in DF200 was important in adapting to the presence of unformylated methionine, while the second mutation in DF200::*def-fmt* day 100 would be a reversion to the wildtype state, and adapting to the presence of formylated methionine once again.

Our *xdhA*. *yecO, yaiT* and *ybaL* mutations did not previously occur in parallel in DF200, making it is difficult to establish its importance in adapting to *def-fmt* loss. As earlier discussed, there could be multiple pathways of adapting to *def-fmt* loss. This means that just because a mutation did not occur in parallel in DF200 lines, it is not necessarily unimportant in adapting to the presence or absence of formylation. Further experiments involving knock-ins of wildtype sequences into evolved *def-fmt* knock-outs (DF200) will allow us to compare the effects of the mutated sequence and the wildtype sequence in the context of a situation where *def-fmt* has been lost, establishing the severity of these mutations. If we observe a loss of fitness upon knocking in this mutation, it is likely that this mutation is important for adapting to the loss of *def-fmt*.

4. Mutations in genes that had been previously mutated, but are not direct reversions

We have also seen further mutations in genes where we have observed previous mutations in DF200, which are not reversions of previous mutations. These are interesting, as they provide potential candidates for adaptation to *def-fmt* reintroduction, which are not direct reversion mutations. This includes *trmB*, where a new C insertion mutation was seen. Previously we have also seen a C insertion in this region which results in a premature stop codon within the protein sequence, and this original mutation occurred upstream of where we have observed the subsequent insertion mutation. What we are observing is consistent with pseudogenisation of a gene that has already lost its function. In our situation, the first mutation in DF200 resulted in a loss of function of this gene, and the subsequent mutation in the DF200::*def-fmt* day 100 lines is this gene starting to pseudogenise. Pseudogenes evolve neutrally as a result of a defective copy of a gene (Khachane & Harrison 2009), and our mutations in DF200::*def-fmt* day 100 could be consistent with pseudogenisation if our mutation in the DF200 line has resulted in inactivation of the protein. Based on the location of our mutation in DF200, and the fact that it has resulted in insertion of a stop codon at residue 88 in the amino acid sequence, we conclude that we have likely lost function of this gene. The function of this gene, and its relevance to our mutation will now be discussed.

trmB was originally denoted as *yggH*, but has since been discovered to act as a tRNA (Guanine-N7)-methyltransferase (Zegers *et al.* 2006), mediating transfer of a methyl group from S-adenosyl-L-methionine to the G46 base in tRNA (Tomikawa *et al.* 2008). Previous experiments on *yggH* in *Escherichia coli* have found this gene to not be essential (De Bie *et al.* 2003). This means that we could theoretically observe inactivating mutations in *trmB*. However, simulations by (Purta *et al.* 2005), have predicted that the important residues for binding activity in tRNA (Guanine-N7)-methyltransferase in *E. coli* are aa 15–40, 148–155, and 214–226. In particular, the E69, E94, and D121 residues are thought to bind methionine, ribose and adenine (referred to as the AdoMet binding site), the main cofactors of this enzyme. Further to this, Purta et al. (2005) performed site directed mutagenesis of regions believed to contribute to the catalytic activity of this enzyme. Three substitution mutants, D144A, R154A and R155A were found to result in a complete loss of activity of the enzyme. With one exception, all other substitutions tested were found to decrease enzymatic activity. As a result of a previous mutation in DF200, we have seen the introduction of a stop codon at residue 83 in the amino acid sequence. The second mutation in DF200::*def-fmt* day 100 at residue 173 in the amino acid sequence has not rescued this mutation, as it occurs considerably downstream of the stop codon introduced by the initial mutation in DF200. This suggests that this protein has been deactivated by mutation, as the first mutation in DF200 has resulted in a stop codon before the region (144-155) considered to be important for protein function (Purta *et al.* 2005). As our computational analyses provide a limited indication as to what is

actually occurring for this protein, experiments focused on the tRNA (Guanine-N7)-methyltransferase will help elucidate whether or not there has been a loss of protein function as a result of our mutations. This might involve site-directed mutagenesis of our wildtype DNA sequence to the mutated sequence, and subsequent analyses of protein function. If we see a loss of protein function as a result of this experiment, we can conclude that our mutations in *trmB* have deactivated the protein, and that our second mutation in DF200::*def-fmt* day 100 lines is the beginning of pseudogenisation of this gene.

We have also seen a further mutation in *infB*, in seven of our eleven DF200::*def-fmt* day 100 lines. Our previous experiments have mapped mutations observed in *infB* to the C2 domain of the IF2 protein in bacteria (Catchpole 2014). This domain is known to be involved with binding of fMet-tRNAi (Sprink *et al.* 2016a). However, due to the low resolution of previous structures, it has been difficult to conclude the effect of observed mutations. We previously mapped observed mutations to a 13.8Å IF2 structure (1ZO1), and while it was possible to hypothesise that these might be changes to adapt to the loss of formylation, it was hard to postulate how these mutations interact with the initiating methionine. More recently, (Sprink *et al.* 2016a) have published a structure of *E. coli* IF2 bound to the ribosome and fMet-tRNAi to a 3.7-Å resolution using cryo-electron microscopy. This structure makes it possible to see how the mutations we have might be interacting with fMet-tRNAi. The structural data suggest further adaptation to the reintroduction of *def-fmt*, with the new S844F mutation

occurring very close to where fMet-tRNAi appears to be binding. From this structure, it would also appear that this mutation is in very close proximity of a purine residue of the fMet-tRNAi, and we have seen a considerable change from a hydrophilic to a hydrophobic residue. Further to this, the S844F mutation is five residues away from a phenalynine residue that appears to be in the proximity of the formyl group on fMet-tRNAi. This suggests that this mutation has a large impact on how IF2 is interacting with fMet-tRNAi, and, as we have seen two mutations (P807S and Y838C) in this region upon knocking out *def-fmt*, we conclude that this further mutation (S844F) occurred in parallel in seven of our lines as a result of *def-fmt* reintroduction. In the future we could use molecular dynamics experiments in order to establish the effect of our mutations on binding to fMet-tRNAi, and thus establish the importance of our mutations to adaptation to *def-fmt* loss and reintroduction. These experiments will be discussed in further detail in Chapter 5.

5. No changes occur within a gene

We have also found that there have been no changes to or ordered loss of the reintroduced *def-fmt* gene pair during the course of our evolution experiment. As previously stated (Chapter 3), if *def-fmt* was spreading through toxin-antitoxin activity alone, this mechanism would not be enough to ensure complete fixation of the system, as ordered loss of the gene pair is expected following genomic integration (Van Melderen & De Bast 2009). However, we have not observed

ordered loss of the *def-fmt* pair in any of our lines (refer to figure 2 and Chapter 3), consistent with a model where this gene pair impacts a core part of translation and spreads quickly so would drive rapid adaptive changes, meaning it is no longer able to be lost.

'Replaying' the initial introduction of *def-fmt* has revealed how dynamic evolutionary processes are. Instead of all the mutations reverting back to the ancestral state, we instead observed a number of different changes, consistent with multiple independent adaptive events following *def-fmt* reintroduction. Such an example of this was seen in IF2. Instead of reversions of previously observed mutations, we see further mutational changes, so that there are now three mutations present in the C2 domain of IF2 in seven of our eleven lines. While we have observed extensive parallelism in our experiments, the variety and scope of mutations we have seen make it clear that not all are predictable. This apparent stochasticity is consistent with theories of evolution previously proposed by Stephen Jay Gould, most particularly, the idea that if we were to replay the tape of life, we would not necessarily reach the same point as before due to the random nature of evolution (Gould 1989). This idea relies very much on the idea of contingency, that every step in evolution is somehow contingent on the one prior, making it difficult to predict the path that evolution will take. We are seeing this idea within some of our mutations, particularly those in *infB*. Instead of reversion back to the ancestral state upon *def-fmt* reintroduction, we are instead seeing further changes in this same gene, showing the idea of contingency in evolution.

However, as a number of our mutational changes are in parallel, this would indicate that, for some mutations and environments the course of evolution is repeatable to some degree.

Chapter 5
Discussion

The use of formylmethionine is conserved across all bacteria. This would indicate that this process is essential for bacterial translation initiation, yet formylmethionine is not used in archaeal or eukaryotic translation. Further to this, knock-outs of *def-fmt* are viable despite a decrease in growth rate, and the function formylation plays is not clear. This raises questions of how this process evolved to be so ubiquitous in bacterial translation. If it is a dispensable process, then how did it evolve to be so well-conserved across bacteria? We have previously shown that formylation, and the subsequent removal of the formyl group by deformylation, has likely evolved as a toxin-antitoxin gene pair present on a selfish genetic element such as a plasmid, and has spread heavily through its capacity for post-segregational killing (Catchpole 2014). This has effectively caused an addiction to the presence of the *def-fmt* gene pair, meaning that it is no longer able to be lost. This was achieved by producing a line devoid of the *def-fmt* gene pair, and evolving replicates of this line for 1,500 generations in the absence of these genes. While these lines initially showed a large decrease in growth rate, after 1,500 generations the growth of these lines was indistinguishable from the wildtype. This ultimately demonstrated that bacteria are capable of wildtype growth in the absence of the *def-fmt* gene pair, showing that a seemingly essential process can be lost from bacteria. Reintroduction of

the *def-fmt* gene pair into these evolved lines on a temperature sensitive plasmid also produced a phenotype comparable to that of a known post-segregational killing system, providing further insight into the evolution of this process (See Chapter 1)(Catchpole 2014).

Addiction to *def-fmt* consistent with toxin-antitoxin activity

In order to further investigate how formylation came to be so heavily conserved in bacteria, this thesis has further investigated evolution of this process. While previous work has shown *def-fmt* capable of acting as a PSK system, and that bacteria are capable of wildtype growth in the absence of *def-fmt* (Catchpole 2014), this work did not investigate how this process evolved to be so universally conserved. We have taken the lines evolved in the absence of *def-fmt* (DF200), and have reintroduced the gene pair (DF200::*def-fmt)*. Immediately we have observed a reassertion of an addiction phenotype, where the gene pair is no longer able to be lost once reintroduced. This was seen in the case of two experiments, both in Chapters 2 and 3. In Chapter 2, we saw addiction to *def*-fmt upon attempts to knock-in *def-fmt.* The methodology used should have theoretically resulted in a 50:50 ratio of the kanamycin-resistance cassette present in the evolved DF200 lines to the reintroduced *def-fmt* gene pair. Instead we have observed solely the *def-fmt* gene pair. Upon treatment with trimethoprim, an antibiotic that prevents the addition of this formyl group (Harvey 1973), and thus formylation, we see the expected 50:50 ratio. This indicates that *def-fmt* is acting to prevent the

emergence of any cells possessing the kanamycin-resistance cassette. Further to this, we have been unable to knock-out the *def-fmt* gene pair at two time points during our evolution experiments. In Chapter 3 we attempted to knock-out the *def-fmt* gene pair at both day 0 and day 80 of our evolution experiment. From the 100 colonies we screened at each time point, none of these cells had lost the *def-fmt* gene pair, once again showing addiction to *def-fmt.* These results, together with results obtained by Catchpole (2014) further support that this system could have evolved through a toxin-antitoxin model, as apparent toxicity appears to being immediately reasserted when the gene pair is interrupted, even when cells have evolved in the absence of the gene pair for 1,500 generations.

Evolutionary contingency observed upon reintroduction of *def-fmt*

Following the further 3,000 generations of culturing we have performed (see Chapter 3), we have sequenced the whole genomes of the eleven lines evolved with the *def-fmt* gene pair reintroduced (Chapter 4). This has revealed a number of compensatory mutations in response to *def-fmt* reintroduction. This includes further mutations in genes where we have already observed changes after *def-fmt* loss, including *infB*, which codes for initiation factor 2, and interacts with fMet-tRNAi^{fMet}. Further to this, we have also seen a number of direct reversion mutations, where we observed mutations in response to the loss of *def-fmt* and have seen a reversion back to the wildtype ancestor state following *def-fmt*

reintroduction. Further to this, we have also observed situations where mutations previously observed in response to *def-fmt* loss have remained unchanged upon reintroduction of the *def-fmt* gene pair. This was seen in a number of cases including the *metZ* promoter which was previously thought to compensate for decreased affinity for unformylated methionine by increasing expression of *metZ.* Finally, we have observed a number of parallel mutations in genes not previously mutated. Some of these have been attributed to adaptation to culturing conditions through their presence in our control lines, with others occurring solely in our experimental lines. These results ultimately show evolution as a dynamic and unpredictable process. We have observed a variety of changes in response to *def-fmt* reintroduction (Chapter 4), and many of these appear to be consistent with the theory of evolutionary contingency, the idea that the evolutionary trajectory of an organism is dependent on contingent evolutionary effects (Gould 1989). This contingency makes it difficult to predict the path that evolution will take, as if we were to replay the tape of life, it would be expected that you would see something entirely different each time. We have observed multiple mutations in the *infB* gene throughout the course of our experiments. Three of these observed mutations appear in the C2 domain of the protein, important for the binding of fMet-tRNAi^{fMet} to initiation factor 2 (IF2). Upon loss of *def-fmt*, two mutations were observed in this region in multiple lines, while a further mutation was observed upon *def-fmt* reintroduction. In this case, instead of seeing a reversion back to the ancestral state of this protein, we are instead seeing further changes, showing the theory of evolutionary contingency and that evolution is

unpredictable. However, as we have seen a number of parallel changes in our experiments, this would indicate that evolution is not entirely unpredictable despite its dynamic nature. The variety and scope of mutational changes we have observed have made it clear that not all evolutionary changes are predictable.

Evolution of formylation is non-selective

The results in this thesis have indicated that the evolution of *def-fmt* is non-adaptive. We have seen a situation where there is no inherent benefit to *def-fmt* being present, yet it is persisting. As we have seen upon reintroduction of the *def-fmt* gene pair, addiction to the genes is occurring immediately (Chapters 2 and 3). While the presence of *def-fmt* is not providing a selective advantage to the cell, it is persisting, as it becomes a disadvantage to lose it once it is present. We are seeing adaptation to its presence once it is reintroduced as a result of this. These results are consistent with the theory of constructive neutral evolution, where even though there is no inherent benefit to a trait, complexity can arise (Gray *et al.* 2010). Under an adaptionist view point, we would expect to see selection for improved function, and that formylation would have persisted through a selective advantage to bacterial translation initiation. However, formylation does not provide any apparent fitness benefits upon introduction (see Chapter 3), and even appears to decrease fitness of the cells very slightly. Constructive neutral evolution instead aims to explain how neutral or even deleterious traits could arise, where such mutations are compensated by further mutations in interacting

constituents (Gray *et al.* 2010). Through this, a complex system might arise through compensatory mutations being more likely than reversions to the ancestral state. The use of formylation in bacterial translation initiation appears to be consistent with such a system. It would appear to be complex, but due to its addictive nature, compensatory mutations are more likely to occur than reversions to the non-formylating ancestral state. We have seen a number of compensatory mutations in response to the reintroduction of *def-fmt*, including mutations in components which interact with the initiating formylmethionine during translation (IF2). This makes constructive neutral evolution a plausible theory for the evolution of formylation in bacterial translation initiation. We are observing a number of mutations that take the evolved lines further from the ancestral state, instead of deactivation or loss of the *def-fmt* gene pair, likely due to the addictive phenotype that it asserts. We are observing a situation where formylation does not necessarily provide an adaptive benefit to cells, but instead one where it becomes beneficial to not lose it once it is present.

Model for the evolution of *def-fmt* in bacterial translation initiation

We have previously proposed a model for the evolution of formylation in bacterial translation initiation (Catchpole 2014). This model states that a non-formylating ancestor to bacteria was invaded by the formylase-deformylase gene

pair on a mobile genetic element such as a plasmid. While initial introduction of *def*-fmt might have been of detriment to translation, it is able to persist due to the post-segregational killing (PSK) action of the gene pair. This gene pair then becomes difficult to lose once it is present, and due to its presence on a plasmid, it is able to spread through horizontal gene transfer. The presence of such a PSK+ plasmid would also provides a situation where other competing plasmids are not able to invade, due to to incompatibility between plasmids (Cooper *et al.* 2010). Over time, we would expect to see reversion to the ancestral state by ordered gene loss, however it is possible for these cells to be reinvaded by the *def-fmt* plasmid. As a result of the PSK activity preventing loss of the plasmid, the ancestor has adapted to the presence of fMet-tRNAi, making it harder to lose the gene pair. Those cells that have adapted to the presence of *def-fmt* would also presumably possess a fitness benefit over those which are not adapted to the presence of *def-fmt* due to its addictive phenotype. At some point during evolutionary history, the gene pair has become integrated into the bacterial chromosome and the plasmid has been lost. Chromsomal integration of *def-fmt* results in a situation where the gene pair is no longer able to be lost. The results we have seen are consistent with this model of *def-fmt* evolution. We have previously seen that *def-fmt* is capable of acting as a PSK system (Catchpole 2014). Now that we have reintroduced *def-fmt* into these lines, we have provided further information consistent with our model. We have seen that *def-fmt* is not able to be lost once it is reintroduced (Chapters 2 and 3), and we have also seen rapid adaptation to its presence (Chapter 4). This would be consistent with our

model, where adaptation occurs due to the inability to lose the gene pair. This model does not disregard any potential proposed functions of formylation. Instead, our model emphasises that formylation likely invaded bacteria as a TA system, and that any subsequent function evolved independent of these evolutionary origins.

We now further add to this model with this thesis. It is likely that horizontal gene transfer of *def-fmt* is occurring following chromosomal integration of the gene pair, as we have observed no ordered loss of the *def-fmt* gene pair in any of our experimental lines. In previous experiments, it has been observed that if a toxin-antitoxin system relies solely on post-segregational killing for its persistence, then selective pressure would eventually result in the loss of the gene-pair due to inactivation of the toxin (Kobayashi 2004). Large numbers of chromosomally-encoded toxin-antitoxin systems have also been observed to be deactivated through inactivation of the toxin (Ramisetty & Santhosh 2016). If *def-fmt* is persisting in our lines solely through its toxin-antitoxin activity, the selection for its maintenance must be extremely strong. In our lines we are seeing no loss of the gene-pair, with an inability to obtain knock-outs once it is reintroduced, and also throughout the course of our long-term evolution experiment. Instead we are seeing a situation where adaptation to *def-fmt* is occurring faster than any potential ordered loss of the genes. Horizontal transfer of *def-fmt* could provide a potential explanation as to why we are not observing such ordered loss of the genes. Recent research has also investigated the possibility that chromosomally-

encoded toxin-antitoxin gene pairs are likely being transferred horizontally between bacteria (Ramisetty & Santhosh 2016). This research found that regions flanking known toxin-antitoxin pairs indicate transfer occurring by non-homologous mechanisms through *de novo* integration. Evidence for homologous mechanisms of integration were seen in *chpBS*, a known toxin-antitoxin gene-pair on the *E. coli* chromosome (Ramisetty & Santhosh 2016). These results, combined with our own results, indicate that it is possible that *def-fmt* is able to be transferred horizontally, and that this could be one of the reasons for its continued persistence in our lines.

Future Directions

We experienced contamination in ten of our eleven control lines, meaning that we were unable to use data from those ten controls. It will be important to complete sequencing of the ten single *def* knock-in lines in order to establish which mutations in our experimental lines are shared with the control lines and which of those are unique to our experimental lines. This will aid in concluding which mutations are important in adapting to *def-fmt* loss and reintroduction. Previously we have focused on mutations we deemed to be important for translation, due to the effect of formylation on translation, and also those mutations which had not occurred in the control lines. In order to avoid bias, it will be important to go back and look at mutations outside of the translational machinery in order to determine any role they might have in adapting to formylation loss or gain. It will also be important to look at the mutations that were also seen within the control lines, as these might also play a role in the increased fitness we observed when adapting to *def-fmt* loss. By investigating other mutations, this will ensure that we are not introducing bias into our research through only investigating those mutations that we deem to be important for the translational apparatus. This will also aid in creating a list of mutations that we determine to be essential for adapting to *def-fmt* loss or reintroduction.

In order to investigate the mutations that are essential for adaptation to loss of *def-fmt* we could also create a clean DF200 line with the minimal number of

mutational changes required for wildtype fitness. To create this line, it will be important to create a list of candidate mutations that are thought to compensate for *def-fmt* loss and reintroduction. This makes it important to resequence our DF200::*def* control lines that were contaminated. Production of a clean DF200 line, however, might be difficult to achieve, as we would have to measure changes in fitness within the DF000 line, a line that grows very slowly. This means that introducing each of these mutations might take an extremely long time, and unless there is a large fitness benefit from a mutation, it might be difficult to see which mutations are of benefit to these lines. Alternatively, we could approach creation of a clean DF200 line by introducing the mutational changes into the REL606 ancestor, and then completing the knock-out of the *def-fmt* gene pair. Providing that the introduced mutations do not result in a loss of fitness, we will be able to perform these introductions far quicker, and then complete the knock-out of the *def-fmt* gene pair. If the introduced mutations provide some inherent benefit to the loss of *def-fmt*, it should theoretically become easier to knock-out the gene pair in this altered line. However, as we have previously shown (see Chapter 3), *def-fmt* appears to have an addictive phenotype immediately upon reintroduction into the DF200 line. This means that the production of a clean DF200 line might be difficult regardless of which approach we take. Instead it might be more logical to establish what mutations are important in our experimental conditions through knocking the wildtype REL606 sequence into our DF200 lines. Any mutations that result in decreased fitness will likely be important in adapting to the loss of *def-fmt.* It might also be possible to

use a computational method to measure the impact of the mutational changes we are seeing. The use of a tool such as delta-bitscore would give an indication of the degree of severity of the mutational changes that we are seeing (Wheeler *et al.* 2016). This technique uses a profile HMM approach, where multiple sequence alignments are used to determine the level of divergence between sequences. Mutations observed in highly conserved regions are likely to be less tolerated than those in non-conserved regions, and would thus give us an indication of the severity of the mutations we are seeing, and whether these are likely to be impacting protein function. This technique will also aid in gaining a greater understanding of the relevance of our mutations to protein function in our lines.

While it has been possible to see bands when completing our northern blot, the small gels we have used have not been able to provide enough of a resolution to distinguish between formylated and unformylated tRNA. This means, that while expression of *def* and *fmt* was confirmed by RT-PCR, it is not possible to tell whether the initiating tRNA is actually being formylated without this data. When running a larger version of the gel for this northern blot, we have encountered difficulties seeing bands. We have attempted a number of remedies in order to solve these issues, including changing of hybridisation temperature, use of DEPC-treated reagents and increasing the amount and concentration of the RNA we have loaded. We have also attempted to change our buffer from a sodium acetate-based solution, to a more conventional TBE buffer. This method appears to have produced the most success, and in the future it will be important to

reattempt this gel with the optimal conditions we have established in order to conclude the state of formylation in the reintroduced lines. If the methods we are currently proposing are not successful, it could be possible to use P32-labelled probes as an alternative to our current usage of DIG-labelled probes in the hope of obtaining a higher resolution. We could also attempt a different method in order to establish whether formylation is once again occurring, such as mass spectrometry, but any technique or approach taken would need to be able to differentiate between formylated and unformylated methionine.

Previously we have investigated the possibility that formylation could be occurring in archaea (Catchpole 2014). Archaeal translation initiation is known to proceed with an unformylated fMet-tRNAi^{Met}, despite evidence of horizontal gene transfer between archaea and bacteria (Aravind *et al.* 1998; Nelson *et al.* 1999; Brochier-Armanet & Forterre 2007). We have previously hypothesised that the absence of formylation in bacterial translation initiation could be due to the presence of deblocking aminopeptidases in archaea (Catchpole 2014), which hydrolyse *N*-blocked amino acids (Ando *et al.* 1999; Story *et al.* 2001; Onoe *et al.* 2002; Jia *et al.* 2011). It was believed that this would result in hydrolysis of formylmethionine, performing the task of peptide deformylase, and therefore preventing any toxic activity of the *def-fmt* gene pair. We sought to investigate whether we see any evidence of *def-fmt* bioinformatically, using the 292 available archaeal genomes at the time (Catchpole 2014). Putative formylase and deformylase genes were found in *Methanocorpusculum labreanum*, and were

predicted to have structural similarity to the bacterial *def-fmt* genes. Further to this, similar *def-fmt*-like sequences were observed in *Methanocorpusculum bavaricum*, and a partial deformylase-like sequence was seen in *Halococcus hamelinensis.* While phylogenetically there is a good signal for formylation, we currently have no experimental evidence to show that this is happening in archaea. In order to test this, we would be able to use the aforementioned northern blot method, using probes specific for the putative archaeal fMet-tRNAi^{Met}, providing the method is established. This information will help to provide a further insight into how formylation is distributed across the domains, and may show that formylation is occurring outside of bacteria, where it has not been previously seen.

Our previous calculations of comparative growth rates are quite crude, having measured the OD595 over a 24-hour period, and calculating doubling time from that (Chapter 2) (Catchpole 2014). In order to look at growth comparisons between different lines, it will be important to perform competition experiments between our lines. Conventionally, an arabinose-marker method has been used on REL606 lines, as it provides an effective method of competing cells (Levin *et al.* 1977; Lenski 1988; Lenski *et al.* 1991b). This method utilises a single nucleotide change in the *araA* gene, where REL606-derived lines are no longer able to utilise arabinose (Maddamsetti *et al.* 2015). Because of this, if these lines are plated on tetrazolium arabinose agar, REL606-derived lines will be pink or red, while REL607-derived lines will be white. This is because utilisation of

arabinose results in excretion of acetic acid, and change in the tetrazolium dye from red to white (Remold & Lenski 2001). This single nucleotide change also makes it easy to select for Ara+ revertants, so that it is possible to compete two REL606-derived strains against each other. However, if the growth difference is small, it can be difficult to compete the cells, and establish exactly which line has the higher level of fitness. As our lines are very close in fitness, it will be difficult to utilise this method. A more recent method that utilises cyan and yellow fluorescent markers could be used in order to compete our cells (Gullberg *et al.* 2011). These markers are variants of green fluorescent protein, that can be visualised using a fluorescence activated cell sorter (FACS), and have been previously inserted into the *galK* gene with a λ red system. We have previously utilised a λ red system when attempting to perform knock-outs of the reintroduced *def-fmt* gene pair (see Chapter 3), so it would be possible to repurpose this system in order to introduce these fluorescent markers. However, in order to use this method, we would first need to establish whether these markers are neutral in our lines. If they are, it will be possible to compete our different lines with these markers, and visualise which cells are outcompeting the others, and therefore the ones with a higher level of fitness. This technique will especially be useful for comparing our lines to the wildtype ancestor, and establishing whether the growth rates we are observing are truly indistinguishable (Catchpole 2014). Similar to this, we could also use a technique such as digital droplet PCR with primers for neutral strain-specific markers in order to quantify relative numbers of lines to establish fitness differences.

While we have previously investigated some aspects of the capacity of the *def-fmt* gene pair to act as a post-segregational killing system, we have not yet successfully investigated whether *def-fmt* is able to be horizontally transferred like known post-segregational killing systems. Many of these systems are known to confer an advantage when the system is present on a plasmid, and is competing against another mobile genetic element (Cooper *et al.* 2010). This is due to the toxin-antitoxin activity, so that a cell is effectively protected from further plasmids being introduced by the presence of a PSK+ plasmid. More simply, a PSK+ plasmid is able to invade cells that contain a PSK- plasmid, but PSK- plasmids are not able to invade cells that contain a PSK+ plasmid. We have previously attempted to replicate an experiment by Cooper *et al.* (2010), but instead testing whether *def-fmt* is able to outcompete invading mobile genetic elements (unpublished). In order to do this, F plasmids were created bearing different antibiotic resistance markers in identical places, and introduced these into the DF200 line. However, initial testing showed that our markers were not neutral. When competing empty vectors against each other, we saw the line with the tetracycline-resistance marker always outcompeting the chloramphenicol-resistant line. This was likely due to our lines suffering from a fitness cost when expressing chloramphenicol resistance. In the future, we could reapproach these experiments by replacing the antibiotic selection marker with something that is neutral, so that the two plasmids do not have any form of fitness effects on our cells. This could be in the form of a different antibiotic marker, or even a different

type of selective marker such as carbohydrate utilisation. We could also forgo the use of selective markers, and instead quantify the ratios of the plasmids using a technique such as q-PCR. This would allow us to compete our cells without any selective marker potentially affecting the cells, and would also take far less time than plating cells in replicate and counting colonies for each plasmid. This experiment could ultimately provide more evidence of the *def-fmt* gene pair acting like known post-segregational killing systems, and potentially help in explaining how formylation evolved to be so ubiquitous.

Finally, we might use molecular dynamics simulation to further investigate our mutations in initiation factor 2 (IF2). While we have established that three of our mutations in IF2 (P807S, Y838C and S844F) are in the C2 domain, the domain considered important for fMet-tRNAi binding (Sprink *et al.* 2016), we cannot definitively conclude how our mutated residues are interacting with the fMet-tRNAi. Molecular dynamics simulations utilise crystallographic, nuclear magnetic resonance (NMR) or homology modelling data (Rakers *et al.* 2015), and allow small-detail simulations (Lindahl 2008). Using Newton's law of motion (incorporating energy potential and forces on individual atoms over time), models of bonded and non-bonded (electrostatic and van der Waals forces) interactions are created (Rakers *et al.* 2015). This methodology allows for analysis of dynamic proteins, and provides information on structure, interaction and movements (Rakers *et al.* 2015). These simulations will aid in elucidating

the relevance of our parallel mutations in IF2 in adapting to *def-fmt* loss and subsequent reintroduction.

To ensure we complete the project in a logical order, it will be important to prioritise our experiments. Paramount will be resequencing the control lines that were initially contaminated. This will ensure that we have a comprehensive list of candidate mutations for a number of other experiments, including any knock-in experiments or computational methods such as Delta-bitscore. Following this, establishing whether there is an addition of a formyl group in our DF200::*def-fmt* lines will be important, as we have not yet confirmed the presence of a formyl group modification in our *def-fmt* knock-in lines (but have confirmed transcripts from the *def-fmt* genes). Finally, other experiments which further delve into the knowledge of formylation, and its effect on translation will be important. This includes experiments on putative archaeal formylase-deformylase genes, as well molecular dynamics studies of our IF2 mutations, and further investigating the capacity of *def-fmt* to act as a toxin-antitoxin system through plasmid competition experiments.

References

Adams JM (1968) On the release of the formyl group from nascent protein. *Journal of Molecular Biology*, **33**, 571–589.

Agrawal RK, Penczek P, Grassucci RA, Li Y, Leith A, Nierhaus, KH, Frank, J (1996) Direct visualization of A-, P-, and E-site transfer RNAs in the *Escherichia coli* Ribosome. *Science*, **271**, 1000–1002.

Anantharaman V, Aravind L (2003) New connections in the prokaryotic toxin-antitoxin network: relationship with the eukaryotic nonsense-mediated RNA decay system. *Genome biology*, **4**, R81–R81.

Ando S, Ishikawa K, Ishida H, Kawarabayasi Y, Kikuchi H, Kosugi Y (1999) Thermostable aminopeptidase from *Pyrococcus horikoshii. FEBS Letters*, **447**, 25–28.

Andreou AZ, Klostermeier D (2014) eIF4B and eIF4G jointly stimulate eIF4A ATPase and unwinding activities by modulation of the eIF4A conformational cycle. *Journal of molecular biology*, **426**, 51–61.

Anfora AT, Welch RA (2006) DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* clinical isolate CFT073. *Journal of Bacteriology*, **188**, 6622–6628.

Antoun A, Pavlov MY, Andersson K, Tenson T, Ehrenberg M (2003) The roles of initiation factor 2 and guanosine triphosphate in initiation of protein synthesis. *The EMBO Journal*, **22**, 5593–5601.

Antoun A, Pavlov MY, Lovmar M, Ehrenberg M (2006) How initiation factors maximize the accuracy of tRNA selection in initiation of bacterial protein synthesis. *Molecular Cell*, **23**, 183–193.

Aravind L, Tatusov RL, Wolf YI, Walker DR, Koonin EV (1998) Evidence for massive gene exchange between archaeal and bacterial hyperthermophiles. *Trends in Genetics*, **14**, 442–444.

Arnez JG, Moras D (1997) Structural and functional considerations of the aminoacylation reaction. *Trends in Biochemical Sciences*, **22**, 211–216.

Banik SD, Nandi N (2012) Mechanism of the activation step of the aminoacylation reaction: a significant difference between class I and class II synthetases. *Journal of Biomolecular Structure and Dynamics*, **30**, 701–715.

Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science*, **289**, 905–920.

Barrick JE, Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews Genetics*, **14**, 827–839.

Barrick JE, Yu DS, Yoon SH Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli. Nature*, **461**, 1243–1247.

Bateman OA, Purkiss AG, van Montfort R Slingsby C, Graham C, Wistow G (2003) Crystal Structure of η-Crystallin: Adaptation of a Class 1 Aldehyde Dehydrogenase for a New Role in the Eye Lens. *Biochemistry*, **42**, 4349–4356.

Becker DM, Guarente L (1991) [12] High-efficiency transformation of yeast by electroporation. *Methods in enzymology*, **194**, 182–187.

Benelli D, Maone E, Londei P (2003) Two different mechanisms for ribosome/mRNA interaction in archaeal translation initiation. *Molecular Microbiology*, **50**, 635–643.

Betzner AS, Keck W (1989) Molecular cloning, overexpression and mapping of the *slt* gene encoding the soluble lytic transglycosylase of *Escherichia coli*. *Molecular & general genetics: MGG*, **219**, 489–491.

Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, Kiefer F, Cassarino TG, Bertoni M, Bordoli L, Schwede T (2014) SWISS-MODEL: modelling protein

tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, **42**, W252–W258.

Bingel-Erlenmeyer R, Kohler R, Kramer G, Sandikci A, Antolic S, Maier T, Schaffitzel C, Wiedmann B, Bukau B, Nenad B (2008) A peptide deformylase-ribosome complex reveals mechanism of nascent chain processing. *Nature*, **452**, 108–111.

Blanco J, Coque JJ, Martin JF (1996) Characterization of the *secA* gene of *Streptomyces lividans* encoding a protein translocase which complements an *Escherichia coli* mutant defective in the ATPase activity of SecA. *Gene*, **176**, 61–65.

Blanquet S, Dessen P, Kahn D (1984) [11] Properties and specificity of methionyl-tRNA fMet formyltransferase from *Escherichia coli*. *Methods in enzymology*, **106**, 141–152.

Bornemann T, Jöckel J, Rodnina MV, Wintermeyer W (2008) Signal sequence–independent membrane targeting of ribosomes containing short nascent peptides within the exit tunnel. *Nature structural & molecular biology*, **15**, 494–499.

Brochier-Armanet C, Forterre P (2007) Widespread distribution of archaeal reverse gyrase in thermophilic bacteria suggests a complex history of vertical inheritance and lateral gene transfers. *Archaea*, **2**, 83–93.

Broom MF, Sherriff RM, Tate WP, Collings J, Chadwick VS (1989) Partial purification and characterization of a formylmethionine deformylase from rat small intestine. *Biochemical Journal*, **257**, 51–56.

Bushnell B (2014) *BBDuk for Geneious 9.0*, http://www.geneious.com/plugins/bbduk

Cantara WA, Murphy FV, Demirci H, Agris PF (2013) Expanded use of sense codons is regulated by modified cytidines in tRNA. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 10964–10969.

Carter AP, Clemons WM, Brodersen DE, Morgan-Warren RJ, Hartsch T, Wimberly BT, Ramakrishnan V (2001) Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science*, **291**, 498–501.

Catchpole RJ (2014) Evolution of The unnecessary: Investigating how fMet became central in bacterial translation initiation. University of Canterbury.

Chang SY, McGary EC, Chang S (1989) Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. *Journal of Bacteriology*, **171**, 4071–4072.

Chattopadhyay MK, Tabor CW, Tabor H (2009) Polyamines are not required for aerobic growth of *Escherichia coli*: Preparation of a strain with deletions in all of the genes for polyamine biosynthesis. *Journal of Bacteriology*, **191**, 5549–5552.

Chen DZ, Patel DV, Hackbarth CJ, Wang W, Dreyer G, Young DC, Margolis PS, Wu C, Ni Z-J, Trias J, White RJ, Yuan Z (2000) Actinonin, a naturally occurring antibacterial agent, is a potent deformylase inhibitor. *Biochemistry*, **39**, 1256–1262.

Chu M, Mierzwa R, He L, Xu L, Gentile F, Terracciano J, Patel M, Miesel L, Bohanon S, Kravec C, Cramer C, Fischman TO, Hruza A, Ramanathan L, Shipkova P, Chan T (2001) Isolation and structure elucidation of two novel deformylase inhibitors produced by *Streptomyces* sp. *Tetrahedron Letters*, **42**, 3549–3551.

Clements JM, Beckett RP, Brown A, Catlin G, Lobell M, Palan S, Thomas W, Whittaker M, Wood S, Salama S (2001) Antibiotic activity and characterization of BB-3497, a novel peptide deformylase inhibitor. *Antimicrobial agents and chemotherapy*, **45**, 563–570.

Coleman J (1992) Characterization of the *Escherichia coli* gene for 1-acyl-sn-glycerol-3-phosphate acyltransferase (*pIsC*). *Molecular and General Genetics MGG*, **232**, 295–303.

Cooper TF, Heinemann JA (2000) Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proceedings of the National Academy of Sciences*, **97**, 12643–12648.

Cooper TF, Heinemann JA (2005) Selection for plasmid post–segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite–level competition. *Proceedings of the Royal Society B: Biological Sciences*, **272**, 403–410.

Cooper TF, Paixão T, Heinemann JA (2010) Within-host competition selects for plasmid-encoded toxin–antitoxin systems. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3149–3155.

Crowell DN, Reznikoff WS, Raetz CR (1987) Nucleotide sequence of the *Escherichia coli* gene for lipid A disaccharide synthase. *Journal of bacteriology*, **169**, 5727–5734.

Cullum AJ, Bennett AF, Lenski RE (2001) Evolutionary adaptation to temperature. IX. Preadaptation to novel stressful environments of *Escherichia coli* adapted to high temperature. *Evolution*, **55**, 2194–2202.

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, **97**, 6640–6645.

De Bast MS, Mine N, Van Melderen L (2008) Chromosomal toxin-antitoxin systems may act as antiaddiction modules. *Journal of bacteriology*, **190**, 4603–4609.

De Bie LGS, Roovers M, Oudjama Y, Wattiez R, Tricot C, Stalon V, Droogmans L, Bujnicki JM (2003) The *yggH* Gene of *Escherichia coli* Encodes a tRNA (m7G46) Methyltransferase. *Journal of Bacteriology*, **185**, 3238–3243.

Dennis PP (1997) Ancient ciphers: translation in archaea. *Cell*, **89**, 1007–1010.

Diago-Navarro E, Hernández-Arriaga AM, Kubik S, Konieczny I, Díaz-Orejas R (2013) Cleavage of the antitoxin of the *parD* toxin–antitoxin system is determined by the ClpAP protease and is modulated by the relative ratio of the toxin and the antitoxin. *Plasmid*, **70**, 78–85.

Dmitriev SE, Stolboushkina EA, Terenin IM, Andreev DE, Garber MB, Shatsky IN (2011) Archaeal translation initiation factor aIF2 can substitute for eukaryotic eIF2 in ribosomal scanning during mammalian 48S complex formation. *Journal of Molecular Biology*, **413**, 106–114.

Duroc Y, Giglione C, Meinnel T (2009) Mutations in three distinct loci cause resistance to peptide deformylase inhibitors in *Bacillus subtilis*. *Antimicrobial agents and chemotherapy*, **53**, 1673–1678.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Ehrenberg M (2010) Protein synthesis: Translocation in slow motion. *Nature*, **466**, 325–326.

Eletsky A, Michalska K, Houliston S, Zhang Q, Daily MD, Xu X, Cui H, Yee A, Lemak A, Wu B, Garcia M, Burnet MC, Meyer KM, Aryal UK, Sanchez O, Ansong C, Xiao R, Acton TB, Adkins JN, Montelione GT, Joachimiak A, Arrowsmith CH, Savchenko A, Szyperski T, Cort JR (2014) Structural and functional characterization of DUF1471 domains of *Salmonella* proteins SrfN, YdgH/SssB, and YahO. *PLoS ONE*, **9**.

Engelberg-Kulka H, Sat B, Reches M, Amitai S, Hazan R (2004) Bacterial programmed cell death systems as targets for antibiotics. *Trends in microbiology*, **12**, 66–71.

Fehér T, Karcagi I, Győrfy Z, Umenhoffer K, Csörgö B, Pósfai G (2008) Scarless engineering of the *Escherichia coli* genome. In: *Microbial Gene Essentiality: Protocols and Bioinformatics*, pp. 251–259. Springer.

Feng S, Chen Y, Gao Y-G (2013) Crystal structure of 70S ribosome with both cognate tRNAs in the E and P sites representing an authentic elongation complex. *PLoS ONE*, **8**, e58829.

Filbin ME, Kieft JS (2009) Toward a structural understanding of IRES RNA function. *Current opinion in structural biology*, **19**, 267–76.

Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer ELL, Tate J, Punta M (2014) Pfam: the protein families database. *Nucleic Acids Research*, **42**, D222–D230.

Gebauer F, Hentze MW (2004) Molecular mechanisms of translational control. *Nature Reviews. Molecular Cell Biology*, **5**, 827–835.

Gerdes K, Christensen SK, Løbner-Olesen A (2005) Prokaryotic toxin–antitoxin stress response loci. *Nature Reviews Microbiology*, **3**, 371–382.

Gerdes K, Gultyaev AP, Franch T, Pedersen K, Mikkelsen ND (1997) Antisense RNA-regulated programmed cell death. *Annual Review of Genetics*, **31**, 1–31.

Gerdes K, Rasmussen PB, Molin S (1986) Unique type of plasmid maintenance function: postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences*, **83**, 3116–3120.

Giglione C, Meinnel T (2001) Organellar peptide deformylases: universality of the N-terminal methionine cleavage mechanism. *Trends in Plant Science*, **6**, 566–572.

Giglione C, Pierre M, Meinnel T (2000a) Peptide deformylase as a target for new generation, broad spectrum antimicrobial agents. *Molecular Microbiology*, **36**, 1197–1205.

Giglione C, Serero A, Pierre M, Boisson B, Meinnel T (2000b) Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. *The EMBO journal*, **19**, 5916–5929.

Gonzalez CF, Proudfoot M, Brown G, Korniyenko Y, Mori H, Savchenko AV, Yakunin AF (2006) Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. *The Journal of Biological Chemistry*, **281**, 14514–14522.

Gordon JJ, Kelly BK, Miller GA (1962) Actinonin: an antibiotic substance produced by an actinomycete. *Nature*, **195**, 701–702.

Gould SJ (1989) *Wonderful life: the Burgess Shale and the nature of history*. W.W. Norton, New York.

Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London B: Biological Sciences*, **205**, 581–598.

Goyal A, Belardinelli R, Maracci C, Milón P, Rodnina MV (2015) Directional transition from initiation to elongation in bacterial translation. *Nucleic Acids Research*, **43**, 10700–10712.

Gray MW, Lukes J, Archibald JM, Keeling PJ, Doolittle W (2010) Irremediable complexity? *Science*, **330**, 920–921.

Grosjean H, Breton M, Sirand-Pugnet P, Tardy F, Thiaucourt F, Citti C, Barré A, Yoshizawa S, Fourmy D, de Crécy-Lagard V (2014) Predicting the minimal translation apparatus: lessons from the reductive evolution of mollicutes. *PLoS genetics*, **10**, e1004363.

Guan Z, Wang X, Raetz CRH (2011) Identification of a chloroform-soluble membrane miniprotein in *Escherichia coli* and its homolog in *Salmonella typhimurium.* *Analytical biochemistry*, **409**, 284–9.

Guillon JM, Heiss S, Soutourina J, Mechulam Y, Laalami S, Grunberg-Manago M, Blanquet S (1996) Interplay of methionine tRNAs with translation elongation factor Tu and

translation initiation factor 2 in *Escherichia coli*. *Journal of Biological Chemistry*, **271**, 22321–22325.

Guillon JM, Mechulam Y, Schmitter JM, Blanquet S, Fayat G (1992) Disruption of the gene for Met-tRNA (fMet) formyltransferase severely impairs growth of *Escherichia coli*. *Journal of bacteriology*, **174**, 4294–4301.

Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI, Lipsitch M (2011) Selection of resistant bacteria at very low antibiotic concentrations. *PLoS Pathogens*, **7**, e1002158.

Hackbarth CJ, Chen DZ, Lewis JG, Clark K, Mangold JB, Cramer JA, Margolic PS, Wang W, Koehn J, Wu C (2002) N-alkyl urea hydroxamic acids as a new class of peptide deformylase inhibitors with antibacterial activity. *Antimicrobial agents and chemotherapy*, **46**, 2752–2764.

Harvey RJ (1973) Growth and initiation of protein synthesis in *Escherichia coli* in the presence of trimethoprim. *Journal of bacteriology*, **114**, 309–22.

Hasenöhrl D, Benelli D, Barbazza A, Londei P, Bläsi U (2006) *Sulfolobus solfataricus* translation initiation factor 1 stimulates translation initiation complex formation. *RNA*, **12**, 674–82.

Hasenöhrl D, Fabbretti A, Londei P, Gualerzi CO, Bläsi U (2009) Translation initiation complex formation in the crenarchaeon *Sulfolobus solfataricus*. *RNA*, **15**, 2288–98.

Hashem Y, des Georges A, Dhote V, Langlois R, Liao HY, Grassucci RA, Hellen CUT, Pestova TV, Frank J (2013) Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*, **153**, 1108–1119.

Helgren TR, Wangtrakuldee P, Staker BL, Hagen TJ (2016) Advances in bacterial methionine aminopeptidase inhibition. *Current Topics in Medicinal Chemistry*, **16**, 397.

Herring CD, Glasner JD, Blattner FR (2003) Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. *Gene*, **311**, 153–163.

Hershberg R, Petrov DA (2008) Selection on codon bias. *Annual Review of Genetics*, **42**, 287–299.

Hinnebusch AG (2006) eIF3: a versatile scaffold for translation initiation complexes. *Trends in biochemical sciences*, **31**, 553–562.

Hirokawa G, Kiel MC, Muto A, Selmer M, Raj VS, Liljas A, Igarashi K, Kaji H, Kaji A (2002) Post-termination complex disassembly by ribosome recycling factor, a functional tRNA mimic. *The EMBO Journal*, **21**, 2272–2281.

Hoeppner MP, Gardner PP, Poole AM (2012) Comparative analysis of RNA families reveals distinct repertoires for each domain of life. *PLoS Computational Biology*, **8**, e1002752.

Holčík M, Iyer VM (1997) Conditionally lethal genes associated with bacterial plasmids. *Microbiology*, **143**, 3403–3416.

Housman D, Jacobs-Lorena M, Rajbhandary UL, Lodish HF (1970) Initiation of haemoglobin synthesis by methionyl-tRNA. *Nature*, **227**, 913–918.

Hsiao C, Mohan S, Kalahar BK, Williams LD (2009) Peeling the onion: ribosomes are ancient molecular fossils. *Molecular biology and evolution*, **26**, 2415–2425.

Ikeuchi Y, Kitahara K, Suzuki T (2008) The RNA acetyltransferase driven by ATP hydrolysis synthesizes N4-acetylcytidine of tRNA anticodon. *The EMBO journal*, **27**, 2194–2203.

Janausch IG, Unden G (1999) The *dcuD* (former *yhcL*) gene product of *Escherichia coli* as a member of the DcuC family of C4-dicarboxylate carriers: lack of evident expression. *Archives of Microbiology*, **172**, 219–226.

Janosi L (1998) Evidence for *invivo* ribosome recycling, the fourth step in protein biosynthesis. *The EMBO Journal*, **17**, 1141–1151.

Jia B, Lee S, Pham BP, Kwack JM, Jin H, Li J, Wang Y, Cheong G-W (2011) Biochemical characterization of deblocking aminopeptidases from the hyperthermophilic archaeon *Thermococcus kodakarensis* KOD1. *Bioscience, Biotechnology, and Biochemistry*, **75**, 1160–1166.

Julián P, Milon P, Agirrezabala X, Lasso G, Gil D, Rodnina MV, Valle M (2011) The cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biology*, **9**, e1001095.

Kaminishi T, Wilson DN, Takemoto C, Harms JM, Kawazoe M, Schluenzen F, Hanawa-Suetsugu K, Shirouzu M, Fucini P, Yokoyama S (2007) A snapshot of the 30S ribosomal subunit capturing mRNA via the Shine-Dalgarno interaction. *Structure*, **15**, 289–297.

Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC (2012) Experimental evolution. *Trends in Ecology & Evolution*, **27**, 547–560.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, **10**, 845–858.

Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, Krummenacker M, Latendresse M, Muñis-Rascado L, Ong Q, Paley S, Schröder I, Shearer AG, Subhraveti P, Travers M, Weerasinghe D, Weiss V, Collado-Vides J, Gunsalus RP, Paulsen I, Karp PD (2013)

EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, **41**, D605–D612.

Khachane AN, Harrison PM (2009) Strong association between pseudogenization mechanisms and gene sequence length. *Biology Direct*, **4**, 1–5.

Kirillov SV, Wower J, Hixson SS, Zimmermann RA (2002) Transit of tRNA through the *Escherichia* coli ribosome: cross-linking of the 3′ end of tRNA to ribosomal proteins at the P and E sites. *FEBS Letters*, **514**, 60–66.

Kobayashi I (2004) Genetic addiction:a principle of gene symbiosis in a genome. In: *Plasmid Biology*. American Society of Microbiology.

Köhrer C, RajBhandary UL (2008) The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. *Methods*, **44**, 129–138.

Koirala S, Wang X, Rao CV (2016) Reciprocal Regulation of L-Arabinose and D-Xylose Metabolism in *Escherichia coli*. *Journal of Bacteriology*, **198**, 386–393.

Kolupaeva VG, Unbehaun A, Lomakin IB, Hellen CUT, Pestova TV (2005) Binding of eukaryotic initiation factor 3 to ribosomal 40S subunits and its role in ribosomal dissociation and anti-association. *RNA*, **11**, 470–486.

Kozak M (1987) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic acids research*, **15**, 8125–8148.

Kozak M (1999) Initiation of translation in prokaryotes and eukaryotes. *Gene*, **234**, 187–208.

Kuczyńska-Wiśnik D, Kedzierska S, Matuszewska E, Lund P, Taylor A. Lipińska B, Laskowska E (2002) The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured in vivo during extreme heat shock. *Microbiology (Reading, England)*, **148**, 1757–1765.

Kuhle B, Ficner R (2014) eIF5B employs a novel domain release mechanism to catalyze ribosomal subunit joining. *The EMBO Journal*, **33**, 1177–1191.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357–359.

Laurberg M, Asahara H, Korostelev A, Zhu J, Trakhanov S, Noller HF (2008) Structural basis for translation termination on the 70S ribosome. *Nature*, **454**, 852–7.

Laursen BS, Siwanowicz I, Larigauderie G, Hedegaard J, Ito K, Nakamura Y, Kenney JM, Mortensen KK, Sperling-Petersen HU (2003) Characterization of mutations in the GTP-binding domain of IF2 resulting in cold-sensitive growth of *Escherichia coli*. *Journal of Molecular Biology*, **326**, 543–551.

Laursen BS, Sørensen HP, Mortensen KK, Sperling-Petersen HU (2005) Initiation of protein synthesis in bacteria. *Microbiology and Molecular Biology Reviews*, **69**, 101–123.

Lee SJ, Jung OS, Lee B-J, Cho K-H, Lee BI (2013) Identification of potent inhibitors against human peptide deformylase as anticancer agents. *Notes*, **34**, 3885.

Lee JH, Pestova TV, Shin B-S, Cao C, Choi SK, Dever TE (2002) Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proceedings of the National Academy of Sciences*, **99**, 16689–16694.

Lee MD, She Y, Soskis MJ, Borella CP, Gardner JR, Hayes PA, Dy BM, Heaney ML, Philips MR, Bornmann WG (2004) Human mitochondrial peptide deformylase, a new anticancer target of actinonin-based antibiotics. *Journal of Clinical Investigation*, **114**, 1107.

Lenski RE (1988) Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution*, 425–432.

Lenski RE (2011) Evolution in action: a 50,000-generation salute to Charles Darwin. *Microbe*, **6**, 30–33.

Lenski RE, Rose MR, Simpson SC, Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist*, **138,** 1315–1341.

Levin BR, Stewart FM, Chao L (1977) Resource-limited growth, competition, and predation: a model and experimental studies with bacteria and bacteriophage. *American Naturalist*, 3–24.

Li X, Chang Y-H (1995) Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proceedings of the National Academy of Sciences*, **92**, 12357–12361.

Li Y, Holmes WB, Appling DR, RajBhandary UL (2000) Initiation of protein synthesis in *Saccharomyces cerevisiae* mitochondria without formylation of the initiator tRNA. *Journal of bacteriology*, **182**, 2886–2892.

Lindahl E (2008) Molecular dynamics simulations. In: *Molecular Modeling of Proteins* Methods Molecular Biology™ (ed Kukol A), pp. 3–23. Humana Press.

Lindner HA, Nadeau G, Matte A, Michel G, Ménard R, Cygler M (2005) Site-directed mutagenesis of the active site region in the Quinate/Shikimate 5-Dehydrogenase YdiB of *Escherichia coli*. *Journal of Biological Chemistry*, **280**, 7162–7169.

Liu Q, Gao Y, Yang W, Zhou H, Gao Y, Zhang X, Teng M, Niu L (2008) Crystallization and preliminary crystallographic analysis of tRNA (m(7)G46) methyltransferase from *Escherichia coli*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, **64**, 743–745.

Livingston DM, Leder P (1969) Deformylation and protein biosynthesis. *Biochemistry*, **8**, 435–443.

Li L, Wang CC (2004) Capped mRNA with a single nucleotide leader is optimally translated in a primitive eukaryote, *Giardia lamblia*. *Journal of Biological Chemistry*, **279**, 14656–14664.

Lokossou AG, Toudic C, Barbeau B (2014) Implication of human endogenous retrovirus envelope proteins in placental functions. *Viruses*, **6**, 4609–4627.

Lomakin IB, Kolupaeva VG, Marintchev A, Wagner G, Pestova TV (2003) Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes & Development*, **17**, 2786–2797.

Londei P (2005) Evolution of translational initiation: new insights from the archaea. *FEMS Microbiology Reviews*, **29**, 185–200.

Lynch M (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 8597–8604.

Machnicka MA, Milanowska K, Osman Oglou O, Purta E, Kurkowska M, Olchowik A, Januszewski W, Kalinowski S, Dunin-Horkawicz S, Rother KM, Helm M, Bujnicki JM, Grosjean H (2013) MODOMICS: a database of RNA modification pathways—2013 update. *Nucleic Acids Research*, **41**, D262–D267.

Maddamsetti R, Lenski RE, Barrick JE (2015) Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli. Genetics*, **200**, 619–10SI.

Margolis P, Hackbarth C, Lopez S, Maniar M, Wang W, Yuan Z, White R, Trias J (2001) Resistance of *Streptococcus pneumoniaeto* deformylase inhibitors is due to mutations in *defB*. *Antimicrobial agents and chemotherapy*, **45**, 2432–2435.

Margolis PS, Hackbarth CJ, Young DC, Wang W, Chen D, Yuan Z, White R, Trias J (2000) Peptide deformylase in *Staphylococcus aureus*: resistance to inhibition is mediated by

mutations in the formyltransferase gene. *Antimicrobial agents and chemotherapy*, **44**, 1825–1831.

Markarian SA, Poladyan AA, Kirakosyan GR, Trchounian AA, Bagramyan KA (2002) Effect of diethylsulphoxide on growth, survival and ion exchange of *Escherichia coli*. *Letters in applied microbiology*, **34**, 417–421.

Mazel D, Coïc E, Blanchard S, Saurin W, Marlière P (1997) A survey of polypeptide deformylase function throughout the eubacterial lineage. *Journal of Molecular Biology*, **266**, 939–949.

Mazel D, Pochet S, Marliere P (1994) Genetic characterization of polypeptide deformylase, a distinctive enzyme of eubacterial translation. *The EMBO journal*, **13**, 914.

Medina V, Pontarollo R, Glaeske D, Tabel H, Goldie H (1990) Sequence of the *pckA* gene of *Escherichia coli* K-12: relevance to genetic and allosteric regulation and homology of *E. coli* phosphoenolpyruvate carboxykinase with the enzymes from *Trypanosoma brucei* and *Saccharomyces cerevisiae*. *Journal of Bacteriology*, **172**, 7151–7156.

Meinnel T, Blanquet S (1993) Evidence that peptide deformylase and methionyl-tRNA (fMet) formyltransferase are encoded within the same operon in *Escherichia coli*. *Journal of bacteriology*, **175**, 7737–7740.

Meinnel T, Blanquet S (1995) Enzymatic properties of *Escherichia coli* peptide deformylase. *Journal of bacteriology*, **177**, 1883–1887.

Melnikov S, Ben-shem A, Garreau De Loubresse N, Jenner L, Yusupova G, Yusupova M (2012) One core, two shells: bacterial and eukaryotic ribosomes. *Nature Structural & Molecular Biology*, **19**, 560–7.

Mengin-Lecreulx D, van Heijenoort J (1996) Characterization of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli*. *The Journal of Biological Chemistry*, **271**, 32–39.

Mi S, Lee X, Xiang-ping L, Veldman GM, Finnerty H, Racie L, LaVaillie E, Tang XY, Edouard P, Howes S, Keith JC, McCoy JM (2000) Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*, **403**, 785–9.

Milligan DL, Koshland DE (1990) The amino terminus of the aspartate chemoreceptor is formylmethionine. *Journal of Biological Chemistry*, **265**, 4455–4460.

Milón P, Rodnina MV (2012) Kinetic control of translation initiation in bacteria. *Critical reviews in biochemistry and molecular biology*, **47**, 334–348.

Mine N, Guglielmini J, Wilbaux M, Van Melderen L (2009) The decay of the chromosomally encoded ccdO157 toxin–antitoxin system in the *Escherichia coli* species. *Genetics*, **181**, 1557–1566.

Mochizuki A, Yahara K, Kobayashi I, Iwasa Y (2006) Genetic addiction: selfish gene's strategy for symbiosis in the genome. *Genetics*, **172**, 1309–1323.

Moll I, Grill S, Gualerzi CO, Bläsi U (2002) Leaderless mRNAs in bacteria: surprises in ribosomal recruitment and translational control. *Molecular Microbiology*, **43**, 239–246.

Moll I, Hirokawa G, Kiel MC, Kaji A, Bläsi U (2004) Translation initiation with 70S ribosomes: an alternative pathway for leaderless mRNAs. *Nucleic acids research*, **32**, 3354–3363.

Naito T, Kusano K, Kobayashi I (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.

Nasvall SJ (2004) The modified wobble nucleoside uridine-5-oxyacetic acid in tRNAProcmo5UGG promotes reading of all four proline codons in vivo. *RNA*, **10**, 1662–1673.

Nasvall SJ, Chen P, Bjork GR (2007) The wobble hypothesis revisited: Uridine-5-oxyacetic acid is critical for reading of G-ending codons. *RNA*, **13**, 2151–2164.

Navarro C, Wu L-F, Mandrand-Berthelot M-A (1993) The *nik* operon of *Escherichia coli* encodes a periplasmic binding-protein-dependent transport system for nickel. *Molecular Microbiology*, **9**, 1181–1191.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, Macdonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, White O, Salzberg SL. Smith HO, Venter JC, Fraser CM (1999) Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323–9.

Newton DT, Creuzenet C, Mangroo D (1999a) Formylation is not essential for initiation of protein synthesis in all eubacteria. *Journal of Biological Chemistry*, **274**, 22143–22146.

Newton DT, Niemkiewicz M, Lo RYC, Mangroo D (1999b) Recognition of the initiator tRNA by the *Pseudomonas aeruginosa* methionyl-tRNA formyltransferase: importance of the base-base mismatch at the end of the acceptor stem. *FEMS microbiology letters*, **178**, 289–298.

Nilsson J, Nissen P (2005) Elongation factors on the ribosome. *Current Opinion in Structural Biology*, **15**, 349–354.

Nilsson AI, Zorzet A, Kanth A, Dahlström S, Berg OG, Andersson DI (2006) Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proceedings of the National Academy of Sciences*, **103**, 6976–6981.

Oberer M, Marintchev A, Wagner G (2005) Structural basis for the enhancement of eIF4A helicase activity by eIF4G. *Genes & development*, **19**, 2212–2223.

O'Donnell SM, Janssen GR (2002) Leaderless mRNAs bind 70S ribosomes more strongly than 30S ribosomal subunits in *Escherichia coli*. *Journal of Bacteriology*, **184**, 6730–6733.

Ogura T, Hiraga S (1983) Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences*, **80**, 4784–4788.

Onoe S, Ando S, Ataka M, Ishikawa K (2002) Active site of deblocking aminopeptidase from *Pyrococcus horikoshii*. *Biochemical and Biophysical Research Communications*, **290**, 994–997.

Otsuka Y (2016) Prokaryotic toxin–antitoxin systems: novel regulations of the toxins. *Current Genetics*, 1–4.

Pahel G, Zelenetz AD, Tyler BM (1978) *gltB* gene and regulation of nitrogen metabolism by glutamine synthetase in *Escherichia coli*. *Journal of Bacteriology*, **133**, 139–148.

Pandey DP, Gerdes K (2005) Toxin–antitoxin loci are highly abundant in free-living but lost from host-associated prokaryotes. *Nucleic acids research*, **33**, 966–976.

Partridge JD, Poole RK, Green J (2007) The *Escherichia coli yhjA* gene, encoding a predicted cytochrome c peroxidase, is regulated by FNR and OxyR. *Microbiology (Reading, England)*, **153**, 1499–1507.

Passmore LA, Schmeing TM, Maag D, Applefield DJ, Acker MG, Algire MA, Lorsch JR, Ramakrishnan V (2007) The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Molecular cell*, **26**, 41–50.

Pauza CD, Karels MJ, Navre M, Schachman HK (1982) Genes encoding *Escherichia coli* aspartate transcarbamoylase: the pyrB-pyrI operon. *Proceedings of the National Academy of Sciences of the United States of America*, **79**, 4020–4024.

Pestova T, Lorsch J, Hellen C (2007) The Mechanism of Translation Initiation in Eukaryotes. In: *Translational control in biology and medicine*, pp. p. 87–128. Hershey JWB: Cold Sring Harbor Laboratory Press.

Petersen HU, Danchin A, Grunberg-Manago M (1976a) Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. I. In vitro studies of the 30S and 70S ribosomal-tRNA complex. *Biochemistry*, **15**, 1357–1362.

Petersen HU, Danchin A, Grunberg-Manago M (1976b) Toward an understanding of the formylation of initiator tRNA methionine in prokaryotic protein synthesis. II. A two-state model for the 70S ribosome. *Biochemistry*, **15**, 1362–1369.

Petersen HU, Joseph E, Ullmann A, Danchin A (1978) Formylation of initiator tRNA methionine in procaryotic protein synthesis: in vivo polarity in lactose operon expression. *Journal of bacteriology*, **135**, 453–459.

Petrov AS, Bernier CR, Hsiao C, Norris AM, Kovacs NA, Waterbury CC, Stepanoc VG, Harvey SC, Fox GE, Wartell RM (2014) Evolution of the ribosome at atomic resolution. *Proceedings of the National Academy of Sciences*, **111**, 10251–10256.

Pfeifer F, Zimmermann P, Scheuch S, Sartorius-Neef S (2005) Gene regulation and the initiation of translation in halophilic archaea. In: *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria, and Eukarya,* Cellular Origin, Life in Extreme Habitats and Astrobiology (eds Gunde-Cimerman N, Oren A, Plemenitaš A), pp. 201–215. Springer Netherlands.

Preiss T (2016) All ribosomes are created equal. Really? *Trends in Biochemical Sciences*, **41**, 121–123.

Preiss T, W. Hentze M (2003) Starting the protein synthesis machine: eukaryotic translation initiation. *BioEssays*, **25**, 1201–1211.

Purta E, van Vliet F, Tricot C, De Bie L, Feder M, Skoqronek K, Droogmans L, Bujnicki JM (2005) Sequence–structure–function relationships of a tRNA (m7G46) methyltransferase studied by homology modeling and site-directed mutagenesis. *Proteins: Structure, Function, and Bioinformatics*, **59**, 482–488.

Putney SD, Sauer RT, Schimmel PR (1981) Purification and properties of alanine tRNA synthetase from *Escherichia coli* A tetramer of identical subunits. *Journal of Biological Chemistry*, **256**, 198–204.

Raina S, Missiakas D, Baird L, Kumar S, Georgopoulos C (1993) Identification and transcriptional analysis of the *Escherichia coli htrE* operon which is homologous to pap and related pilin operons. *Journal of Bacteriology*, **175**, 5009–5021.

Rajagopalan PTR, Datta A, Pei D (1997) Purification, characterization, and inhibition of peptide deformylase from *Escherichia coli*†. *Biochemistry*, **36**, 13910–13918.

Rakers C, Bermudez M, Keller BG, Mortier J, Wolber G (2015) Computational close up on protein–protein interactions: how to unravel the invisible using molecular dynamics simulations? *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **5**, 345–359.

Ramachandran R, Hartmann C, Song HK, Huber R, Bochtler M (2002) Functional interactions of HslV (ClpQ) with the ATPase HslU (ClpY). *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7396–7401.

Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. *Cell*, **108**, 557–572.

Ramesh V, Köhrer C, RajBhandary UL (2002) Expression of *Escherichia coli* methionyl-tRNA formyltransferase in *Saccharomyces cerevisiae* leads to formylation of the cytoplasmic initiator tRNA and possibly to initiation of protein synthesis with formylmethionine. *Molecular and cellular biology*, **22**, 5434–5442.

Ramisetty BCM, Santhosh RS (2016) Horizontal gene transfer of chromosomal Type II toxin–antitoxin systems of Escherichia coli. *FEMS Microbiology Letters*, **363**, fnv238.

Rankin DJ, Turner LA, Heinemann JA, Brown SP (2012) The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3706–3715.

Rapp M, Granseth E, Seppälä S, von Heijne G (2006) Identification and evolution of dual-topology membrane proteins. *Nature Structural & Molecular Biology*, **13**, 112–116.

Reidl J, Römisch K, Ehrmann M, Boos W (1989) MalI, a novel protein involved in regulation of the maltose system of *Escherichia coli*, is highly homologous to the repressor proteins GalR, CytR, and LacI. *Journal of Bacteriology*, **171**, 4888–4899.

Remold SK, Lenski RE (2001) Contribution of individual random mutations to genotype-by-environment interactions in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, **98**, 11388–11393.

Rice KC, Bayles KW (2003) Death's toolbox: examining the molecular components of bacterial programmed cell death. *Molecular microbiology*, **50**, 729–738.

Rodnina MV, Beringer M, Wintermeyer W (2007) How ribosomes make peptide bonds. *Trends in Biochemical Sciences*, **32**, 20–26.

Rogers GW, Richter NJ, Lima WF, Merrick WC (2001) Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F. *Journal of Biological Chemistry*, **276**, 30914–30922.

Rudd KE (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Research*, **28**, 60–64.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning*. Cold spring harbor laboratory press New York.

Samuel CE, D'Ari L, Rabinowitz JC (1970) Evidence against the folate-mediated formylation of formyl-accepting methionyl transfer ribonucleic acid in *Streptococcus faecalis* R. *Journal of Biological Chemistry*, **245**, 5115–5121.

Samuel CE, Rabinowitz JC (1974) Initiation of protein synthesis by folate-sufficient and folate-deficient *Streptococcus faecalis* R: Biochemical and biophysical properties of methionine transfer ribonucleic acid. *Journal of Biological Chemistry*, **249**, 1198–1206.

Sartorius-Neef S, Pfeifer F (2004) In vivo studies on putative Shine–Dalgarno sequences of the halophilic archaeon *Halobacterium salinarum*. *Molecular microbiology*, **51**, 579–588.

Schluenzen F, Tocilj A, Zarivach R, Harms J, Gluehmann M, Janell D, Bashan A, Bartels H, Agmon I, Franceschi F, Yonath A (2000) Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell*, **102**, 615–623.

Schmeing TM, Ramakrishnan V (2009) What recent ribosome structures have revealed about the mechanism of translation. *Nature*, **461**, 1234–1242.

Schmitt E, Blanquet S, Mechulam Y (1999) Crystallization and preliminary X-ray analysis of *Escherichia coli* methionyl-tRNAMetf formyltransferase complexed with formyl-methionyl-tRNAMetf. *Acta Crystallographica Section D: Biological Crystallography*, **55**, 332–334.

Schmitt ME, Brown TA, Trumpower BL (1990) A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae. Nucleic acids research*, **18**, 3091.

Schomburg D, Schomburg I, Chang A (Eds.) (2006) Methionyl-tRNA formyltransferase. In: *Springer Handbook of Enzymes* Springer Handbook of Enzymes. Springer Berlin Heidelberg, Berlin, Heidelberg.

Schrödinger, LLC (2015) The PyMOL Molecular Graphics System, Version 1.8.

Schultz JE, Schulz JE, Matin A (1991) Molecular and functional characterization of a carbon starvation gene of *Escherichia coli*. *Journal of Molecular Biology*, **218**, 129–140.

Sergiev PV, Lesnyak DV, Kiparisov SV, Burakovsky DE, Leonov AA, Bogdanov AA, Brimacombe R, Dontsova OA (2005) Function of the ribosomal E-site: a mutagenesis study. *Nucleic Acids Research*, **33**, 6048–6056.

Simonetti A, Marzi S, Myasnikov AG, Fabbretti A, Yusupov M. Gualerzi CO, Klaholz BP (2008) Structure of the 30S translation initiation complex. *Nature*, **455**, 416–420.

Spiro S, Roberts R., Guest J. (1989) FNR-dependent repression of the ndh gene of *Escherichia coli* and metal ion requirement for FNR-regulated gene expression. *Molecular Microbiology*, **3**, 601 – 608.

Sprink T, Ramrath DJF, Yamamoto H*,* Yamamoto K, Loerke J, Ismer J, Hildebrand PW, Scheerer P, Bürger J, Mielke T, Spahn CMT (2016) Structures of ribosome-bound initiation factor 2 reveal the mechanism of subunit association. *Science Advances*, **2**, e1501502.

Stark H, Rodnina MV, Wieden H-J, Zemlin F, Wintermeye W, Heel M (2002) Ribosome interactions of aminoacyl-tRNA and elongation factor Tu in the codon-recognition complex. *Nature Structural Biology*, **9**, 849–854.

Steiner-Mosonyi M, Creuzenet C, Keates RAB, Strub BR, Mangroo D (2004) The *Pseudomonas aeruginosa* initiation factor IF-2 is responsible for formylation-independent protein initiation in *P. aeruginosa*. *Journal of Biological Chemistry*, **279**, 52262–52269.

Stern L, Schulman L (1978) The role of the minor base N4-acetylcytidine in the function of the *Escherichia coli* noninitiator methionine transfer RNA. *Journal of Biological Chemistry*, **253**, 6132–6139.

Stoltzfus A (1999) On the possibility of constructive neutral evolution. *Journal of Molecular Evolution*, **49**, 169–181.

Story SV, Grunden AM, Adams MWW (2001) Characterization of an aminoacylase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Journal of Bacteriology*, **183**, 4259–4268.

Terenin IM, Akulich KA, Andreev DE, Polyanskaya SA, Shatsky IN, Dmitriev SE (2016) Sliding of a 43S ribosomal complex from the recognized AUG codon triggered by a delay in eIF2-bound GTP hydrolysis. *Nucleic Acids Research*, **44**, 1882–1893.

The UniProt Consortium (2015) UniProt: a hub for protein information. *Nucleic Acids Research*, **43**, D204–D212.

Thoma C, Fraterman S, Gentzel M, Wilm M, Hentze MW (2008) Translation initiation by the c-myc mRNA internal ribosome entry sequence and the poly(A) tail. *RNA*, **14**, 1579–1589.

Tomikawa C, Ochi A, Hori H (2008) The C-terminal region of thermophilic tRNA (m7G46) methyltransferase (TrmB) stabilizes the dimer structure and enhances fidelity of methylation. *Proteins: Structure, Function, and Bioinformatics*, **71**, 1400–1408.

Trobro S, Åqvist J (2005) Mechanism of peptide bond synthesis on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12395–12400.

Valle M, Zavialov A, Li W, Stagg SM, Sengupta J, Nielsen RC, Nissen P, Harvey SC, Ehrenberg M, Frank J (2003) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Structural Biology*, **10**, 899–906.

Van Melderen L, De Bast MS (2009) Bacterial toxin–antitoxin systems: more than selfish entities? *PLoS genetics*, **5**, e1000437.

Villa E, Sengupta J, Trabuco LG, LeBarron J, Baxter WT, Shaikh TR, Grassucci RA, Nissen P, Ehrenberg M, Schulten K, Frank J (2009) Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 1063–1068.

Voloshin ON, Camerini-Otero RD (2007) The DinG protein from *Escherichia coli* is a structure-specific helicase. *Journal of Biological Chemistry*, **282**, 18437–18447.

Weingarten-Gabbay S, Elias-Kirma S, Nir R, Gritsenko AA, Stern-Ginossar N, Yakhini Z, Weinberger A, Segal E (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science*, **351**, 240–240.

Weisser M, Voigts-Hoffmann F, Rabl J, Leibundgut M, Ban N (2013) The crystal structure of the eukaryotic 40S ribosomal subunit in complex with eIF1 and eIF1A. *Nature structural & molecular biology*, **20**, 1015–1017.

Wells SE, Hillner PE, Vale RD, Sachs AB (1998) Circularization of mRNA by Eukaryotic Translation Initiation Factors. *Molecular Cell*, **2**, 135–140.

Wheeler NE, Barquist L, Kingsley RA, Gardner PP (2016) A profile-based method for identifying functional divergence of orthologous genes in bacterial genomes. *bioRxiv*, 022616.

Wienk H, Tishchenko E, Belardinelli R, Tomaselli S, Dongre R, Spurio R, Folkers GE, Gualerzi CO, Boelens R (2012) Structural dynamics of bacterial translation initiation factor IF2. *The Journal of Biological Chemistry*, **287**, 10922–10932.

Wiser MJ, Ribeck N, Lenski RE (2013) Long-term dynamics of adaptation in asexual populations. *Science*, **342**, 1364–1367.

Wurtzel O, Sapra R, Chen F, Zhu Y, Simmons BA, Sorek R (2010) A single-base resolution map of an archaeal transcriptome. *Genome research*, **20**, 133–141.

Yamaguchi Y, Park J-H, Inouye M (2011) Toxin-Antitoxin systems in bacteria and archaea. *Annual Review of Genetics*, **45**, 61–79.

Yuan Z, Trias J, White RJ (2001) Deformylase as a novel antibacterial target. *Drug Discovery Today*, **6**, 954–961.

Yuan Z, White RJ (2006) The evolution of peptide deformylase as a target: Contribution of biochemistry, genetics and genomics. *Biochemical Pharmacology*, **71**, 1042–1047.

Zegers I, Gigot D, van Vliet F, Tricot C, Aymerich S, Bujnicki K, Kosinski J, Droogmans L (2006) Crystal structure of *Bacillus subtilis* TrmB, the tRNA (m7G46) methyltransferase. *Nucleic Acids Research*, **34**, 1925–1934.

Zielenkiewicz U, Cegłowski P (2005) The toxin-antitoxin system of the streptococcal plasmid pSM19035. *Journal of bacteriology*, **187**, 6094–6105.